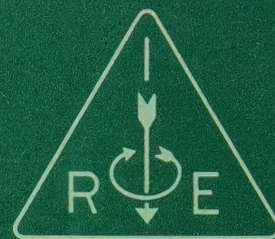


# IRE Transactions



on INFORMATION THEORY

Volume IT-2

DECEMBER, 1956

Number 4

## In This Issue

PERIODICAL  
UNIVERSITY OF HAWAII  
LIBRARY

Frontispiece	page 100
Editorial	page 101
Contributions	page 102
PGIT News	page 154
Contributors	page 155
Annual Index 1956	follows page 156

For complete Table of Contents, see page 99.

PUBLISHED BY THE  
Professional Group on Information Theory

## IRE PROFESSIONAL GROUP ON INFORMATION THEORY

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of the prescribed annual assessment of \$2.00.

### ADMINISTRATIVE COMMITTEE

Chairman	Vice-Chairman	Secretary-Treasurer
M. J. DiTORO Polytechnic Research and Development Co., Inc. Brooklyn 1, New York	WILBUR B. DAVENPORT, JR. Lincoln Laboratories Massachusetts Institute of Technology Cambridge 39, Massachusetts	SID DEUTSCH Microwave Research Institute Brooklyn 1, New York
T. P. CHEATHAM Melpar, Inc. Boston, Mass.	M. J. E. GOLAY Ridge Road and Auldwood Lane, Rumson, N. J.	
HAROLD DAVIS Department of Engineering University of California Los Angeles 14, California	HAROLD R. HOLLOWAY American Machine and Foundry Co., Greenwich, Connecticut	
LOUIS A. DE ROSA Federal Telecommunication Laboratories, Nutley, New Jersey	ERNEST R. KRETZMER Bell Telephone Laboratories Murray Hill, New Jersey	
DONAL B. DUNCAN North American Aviation Downey, California	NATHAN MARCHAND Marchand Electronic Laboratories Greenwich, Connecticut	
R. M. FANO Research Laboratory of Electronics Massachusetts Institute of Technology Cambridge 39, Massachusetts	WINSLOW PALMER Sperry Gyroscope Company Great Neck, New York	
LAURIN G. FISCHER Federal Telecommunication Labs., Nutley, New Jersey	F. L. H. M. STUMPERS N. V. Philips Gloeilampfabrieken Research Labs., Eindhoven, Netherlands	
	W. D. WHITE Airborne Instruments Laboratory, Inc. 160 Old Country Road Mineola, New York	

### IRE TRANSACTIONS®

#### on Information Theory

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Information Theory at 1 East 79th Street, New York 21, N. Y. Responsibility for the contents rests upon the authors, and not upon the Institute, the Group or its members. Single copy prices: IRE-PGIT members, \$1.85; IRE members, \$2.75; nonmembers, \$5.55.

Copyright © 1956—THE INSTITUTE OF RADIO ENGINEERS, INC.

All rights, including translation, are reserved by the IRE. Requests for republication privileges should be addressed to the Institute of Radio Engineers, 1 E. 79th St., New York 21, N. Y.

# IRE Transactions

## on

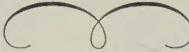
# Information Theory

*Published Quarterly by the Professional Group on Information Theory*


Volume IT-2

December, 1956

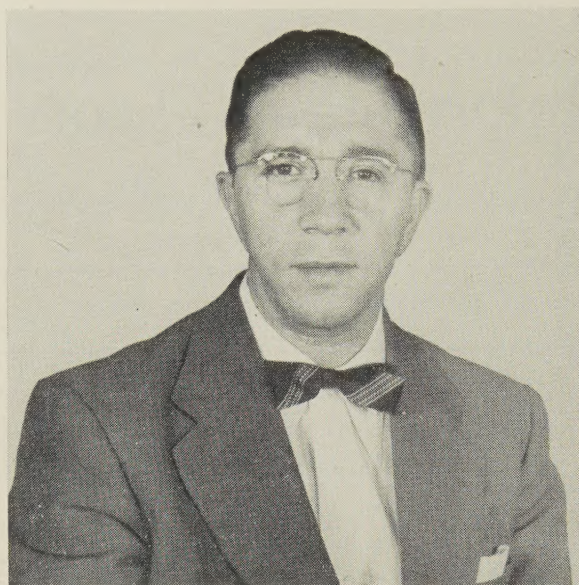
Number 4



### TABLE OF CONTENTS



	PAGE
Frontispiece	<i>Michael J. Di Toro, Jr.</i> 100
Editorial	<i>Michael J. Di Toro, Jr.</i> 101
Contributions	
On the Shannon Theory of Information Transmission in the Case of Continuous Signals	<i>Andrei N. Kolmogorov</i> 102
On Noise Stability of a System with Error-Correcting Codes	<i>Vladimir I. Siforov</i> 109
Two Inequalities Implied by Unique Decipherability	<i>Brockway McMillan</i> 115
A Note on the Maximum Flow Through a Network	<i>P. Elias, A. Feinstein, and C. E. Shannon</i> 117
Rectification of Two Signals in Random Noise	<i>L. Lorne Campbell</i> 119
Optimum Detection of Random Signals in Noise, with Application to Scatter-Multi-path Communication, I	<i>Robert Price</i> 125
A Coincidence Procedure for Signal Detection	<i>Mischa Schwartz</i> 135
Some General Aspects of the Sampling Theorem	<i>D. L. Jagerman and L. J. Fogel</i> 139
The Axis-Crossing Intervals of Random Functions	<i>J. A. McFadden</i> 146
Determination of Redundancies in a Set of Patterns	<i>Arthur Glovazky</i> 151
Correction to "Optimum, Linear, Discrete Filtering of Signals Containing a Nonrandom Component"	<i>Kent R. Johnson</i> 154
PGIT News	
Report on 1956 Symposium on Information Theory	154
Contributors	154
Annual Index 1956	<i>follows page</i> 156



**MICHAEL J. DI TORO, JR.**

Michael J. Di Toro was born in Campobasso, Italy, on June 24, 1910 and came to the United States in 1916. He received the E.E. degree in 1931, the M.E.E. degree in 1933, and the D.E.E. degree in 1946, all from the Polytechnic Institute of Brooklyn.

In 1934, Dr. Di Toro joined the laboratories of the Thomas A. Edison Company in West Orange, N.J. where he engaged in the development of electro-acoustical and electromechanical devices used in recording and reproducing sound. In 1938, while in charge of acoustical development, he published the first analytical paper on phonograph tracing distortion.

He became associated with the Hazeltine Electronics Corporation in 1941, and, as senior electrical engineer, was in charge of projects concerned with air-ground pulse time modulated telemetering, vhf wave meters, and the development and production of miniaturized large bandwidth-delay product video delay lines, for which he evolved the new technique of multilayer winding with self capacitance for the correction of phase distortion.

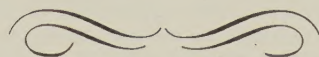
In 1946 he joined the Microwave Research Institute of Polytechnic Institute of Brooklyn, becoming assistant director in 1947. He supervised and participated in projects on uhf power meters, fm transient distortion analysis, and nonlinear network research.

At present he is an adjunct professor in the graduate school of electrical engineering

From 1947 to 1952 Dr. Di Toro was with Federal Telecommunications Laboratories where he initiated and supervised several projects for the armed forces. These included a new type scanning comb filter receiver for passive sonar detection, speech compression and noise suppression systems, and reliable long-range pulsed ionospheric air-ground radio communication system.

During 1952-1953 he was in charge of a missile instrumentation project for A.B. DuMont Laboratories in Passaic, N.J. In 1953 he accepted the position of head of the Electronics Development Division at Fairchild Guided Missile Division in Wyandanch, L.I. He joined the Polytechnic Research and Development Co., Inc. in Brooklyn, N.Y., as chief electronics engineer in 1955, and became Director of Engineering in 1956.

Dr. Di Toro was elected a Fellow of the American Acoustical Society in 1953. He is a senior member of the IRE, and chairman of the PGIT. He is also a member of the AIEE, the American Physical Society, the American Management Association, Sigma Xi, and Eta Kappa Nu, and is a registered professional engineer in New York. He has received twenty patents.



# Applications of Information Theory

MICHAEL J. DI TORO, JR.

Chairman, Professional Group on Information Theory

It is now almost four years since the appearance of the first IRE TRANSACTIONS ON INFORMATION THEORY, and even longer since the pioneering work in this field of Wiener, Shannon, Fano, Woodward, and others. A large number of papers have been published and already six symposia of international scope have been held in London, Boston, and New York. Besides the direct communication, radar and data handling applications, other diverse fields have been considered such as machine translation, reliable computer operation, automatic speech recognition, pattern learning and recognition, theory of hearing, management, and many others. It cannot be doubted that, with the continual expenditure of sufficient funds and manpower, the cross-fertilization arising from all this activity should increase the rate of conception of new ideas.

But conception is wonderful—then comes labor. Has not the time been reached when it is necessary to take stock of progress in knowledge since the beginning of the era of Information Theory? Should not we determine now how such learning may be applied, and with what expected payoff, in the design of economically justifiable equipment in the wide fields of interest of the other twenty-three professional groups of the IRE?

An example of what has been done without aid from Information Theory is the discovery of pulse code modulation by A. H. Reeves in 1939, representing a major breakthrough in a method of coding. It is quite remarkable that pcm comes within about 8 db of signal/noise (for a practically small error probability) to the ideal channel capacity given by Shannon's now famous law. On the other hand, even with all of the available body of theory, consider the rather meager results relative to expenditure of funds and effort obtained in some of the bizarre schemes proposed to conserve bandwidth in speech and tv or fax signals, when viewed in the sobering light of economic justification of the ensuing equipment.

It is obviously necessary to encourage the application of the teachings of Information Theory by engineers in the other IRE professional groups. This can be aided considerably by PGIT through publication of original and higher

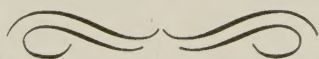
caliber papers which can be written in the beautiful and practical style exemplified by J. R. Carson's papers. Thus, it would seem desirable that the problem considered by an author should be stated preferably in nonmathematical language, and followed by a review of prior work and its limitations. The new results of the paper may then be given together with some typical nontrivial applications, and finally, in an appendix which need not be read nor understood by the reader in order for him to apply the results of the paper, the usual pyrotechnical display of the author's mathematical prowess may be given.

Of particular importance are papers stating basic laws of Information Theory which set theoretical bounds on performance beyond which one cannot go and therefore should not spend money and effort attempting to do so. Examples are Shannon's channel capacity law, Wiener's irreducible residual error in least-mean-square-error linear filtering, Gabor's uncertainty principle between frequency uncertainty and time of occurrence uncertainty, and in the related field of circuit theory, the Paley-Wiener theorem regarding the nonphysical realizability of such structures as the ideal filter with rectangular amplitude and linear phase responses. These are the important tools by which, in the absence of a unified theory of system synthesis, one can be guided in the design and evaluation of suggested systems.

Some of the future channels of research in Information Theory which need further investigation have been recently discussed briefly by Fano in a May, 1956 report for the Technical Advisory Panel on Electronics, Office of the Assistant Secretary of Defense. It is urged that interested readers who are in a position to stimulate activity in the indicated areas should do so.

Moreover it is also stressed that consideration be given to the man-machine matching problem, wherein the machine enhances, rather than replaces, man's ear-brain, eye-brain, and his other highly developed functions which cannot now be satisfactorily replaced by the machine.

Again, as in previous editorials, I should like to encourage PGIT members to air their views (*e.g.*, take exceptions to the foregoing) in our PGIT Correspondence section.



In addition to the scheduled program, the following two papers, by A. N. Kolmogorov and V. I. Siforov, were presented at the 1956 Symposium on Information Theory. However, the manuscripts were received too late for inclusion in the September (Symposium) issue of these TRANSACTIONS. The papers were submitted in response to our invitation to these distinguished Russian scientists, and the following translations were distributed to those attending the Symposium.—*The Editor*.

# On the Shannon Theory of Information Transmission in the Case of Continuous Signals\*

ANDREI N. KOLMOGOROV†

## I. INTRODUCTION

THE ROLE of the entropy of a random object  $\xi$ , capable of taking the values  $x_1, x_2, \dots, x_n$  with the probabilities  $p_1, p_2, \dots, p_n$ ,

$$H(\xi) = - \sum_k p_k \log p_k$$

in information theory and in the theory of information transmission using discrete signals, can be considered to have been explained sufficiently. Furthermore, I insist that the fundamental concept, which admits of generalization to perfectly arbitrary continuous information and signals, is not directly the entropy concept but the concept of the quantity of information  $I(\xi, \eta)$  in the random object  $\xi$  relative to the object  $\eta$ . In the discrete case this quantity is evaluated correctly according to the well-known Shannon formula:<sup>1</sup>

$$I(\xi, \eta) = H(\eta) - MH(\eta/\xi).$$

For a finite-dimensional distribution, possessing density, the quantity  $I(\xi, \eta)$  is determined, according to Shannon, by the analogous formula

$$I(\xi, \eta) = h(\eta) - Mh(\eta/\xi),$$

where  $h(\eta)$  is the "differential entropy"

$$h(\eta) = - \int p(y) \log p(y) dy,$$

and  $h(\eta/\xi)$  is the conditional differential entropy defined in an analogous manner. It is well known that the quantity  $h(\xi)$  has no direct real interpretation and is not even invariant with respect to coordinate transformation in the space of the  $x$ 's. For an infinitely-dimensional distribution, the analog of  $h(\xi)$  is nonexistent, in general.

According to the proper meaning of the word, the entropy of the object  $\xi$  with a continuous distribution is

always infinite. If the continuous signals can, nevertheless, serve to transmit finitely great information, then it is only because they are always observed with bounded accuracy. Consequently, it is natural to define the appropriate " $\epsilon$ -entropy"  $H_\epsilon(\xi)$  of the object  $\xi$  by giving the accuracy of observation  $\epsilon$ . Shannon did thus under the designation "rate of creating information with respect to a fidelity criterion." Although choosing a new name for this quantity does not alter the situation, I decided to call your attention to that proposition by underlining the more widespread interest in the concept and its deep analogy to the ordinary exact entropy. I imply, beforehand, that, as remarked in Section IV, the theorem on the extremal role of the normal distribution (both in the finite-dimensional and the infinite-dimensional cases) is retained for the  $\epsilon$  entropy. Furthermore, I give in Sections II and III, without pretending to its unconditional newness, an abstract formulation of the definition and fundamental properties of  $I(\xi, \eta)$  and a survey of the fundamental problems of the Shannon theory of information transmission. Certain specific results obtained recently by Soviet investigators are explained in Sections IV to VI. I wish to emphasize, especially, the very significant interest, as it appears to me, in the investigations of the asymptotic behavior of the  $\epsilon$  entropy as  $\epsilon \rightarrow 0$ . The cases investigated earlier

$$H_\epsilon(\xi) \sim n \log \frac{1}{\epsilon}; \quad \overline{H}_\epsilon(\xi) = 2w$$

where  $n$  is the number of measurements and  $w$  is the bandwidth of the spectrum, are only very particular cases of the rules which can be encountered here. In order to understand the perspectives disclosed here, my note,<sup>2</sup> explained in another terminology, might be of interest; hence, I am placing a certain number of reprints at the disposal of the participants of the symposium.

To a considerable degree, my report reproduces the contents of a report presented jointly with Iaglom and Gel'fand at the Third All-Union Mathematics Conference.

\* Presented at 1956 Symposium on Information Theory at Mass. Inst. Tech., Cambridge, Mass., September 10-12, 1956. Translated by Morris D. Friedman.

† Academician, Academy of Science, USSR.

<sup>1</sup> It seems expedient to me that the notation  $H(\eta/x)$  is the conditional entropy of  $\eta$  for  $\xi = x$  and  $MH(\eta/\xi)$  is the mathematical expectation of this conditional entropy for the variable  $\xi$ .

<sup>2</sup> A. N. Kolmogorov, *Doklady, AN USSR*, vol. 108, no. 3, pp. 385-388; 1956.

However, since the present symposium is of a more engineering character, I omitted a number of mathematical details. The work of Khinchin on the logical foundations of the theory remains beyond the limits of my survey.

As regards the work of Soviet radio engineers, you will hear about some of them from the other speakers. In the note itself, I will have occasion to note only the interest, in principle, of certain early work of Kotel'nikov, circa 1933 (see Section VI, further).

## II. QUANTITY OF INFORMATION IN ONE RANDOM OBJECT RELATIVE TO ANOTHER

Let  $\xi$  and  $\eta$  be random objects with regions of possible values  $X$  and  $Y$ ,

$$P_{\xi}(A) = P(\xi \in A); \quad P_{\eta}(B) = P(\eta \in B)$$

the appropriate probability distributions, and

$$P_{\xi\eta}(C) = P((\xi, \eta) \in C)$$

the joint probability distribution of the objects  $\xi$  and  $\eta$ . By definition, the quantity of information in the random object  $\xi$  relative to the random object  $\eta$  is given by the formula

$$I(\xi, \eta) = \int_X \int_Y P_{\xi\eta}(dx dy) \log \frac{P_{\xi\eta}(dx dy)}{P_{\xi}(dx)P_{\eta}(dy)}. \quad (1)$$

The exact meaning of this formula requires certain elucidation and the general properties of  $I(\xi, \eta)$ , given later, are correct only for certain limitations of a set-theoretical character on the distributions  $P_{\xi}$ ,  $P_{\eta}$  and  $P_{\xi\eta}$ , but I will not dwell on this here. In every case, the general theory can be explained, without great difficulty, in such a way that it will be applicable to random objects  $\xi$  and  $\eta$  of every general nature (vectors, functions, generalized functions, etc.).

Eq. (1) can be considered to be due to Shannon although he was limited to the case

$$P_{\xi}(A) = \int_A p_{\xi}(x) dx; \quad P_{\eta}(B) = \int_B p_{\eta}(y) dy$$

$$P_{\xi\eta}(C) = \iint_C p_{\xi\eta}(x, y) dx dy$$

when (1) transforms into

$$I(\xi, \eta) = \int_X \int_Y P_{\xi\eta}(x, y) \log \frac{P_{\xi\eta}(x, y)}{P_{\xi}(x)P_{\eta}(y)} dx dy.$$

Sometimes, it is useful to represent the distribution  $P$  as

$$P_{\xi\eta}(C) = \iint_C a(x, y) P_{\xi}(dx) P_{\eta}(dy) + S(C) \quad (2)$$

where the function  $S(C)$  is singular relative to the product

$$P_{\xi} \times P_{\eta}.$$

If the singular component of  $S$  is lacking, then the formula

$$\alpha_{\xi\eta} = a(\xi, \eta) \quad (3)$$

determines the random quantity  $\alpha_{\xi\eta}$  uniquely to the accuracy of probability zero. Sometimes, the following theorem formulated by Gel'fand and Iaglom<sup>3</sup> is useful.

Theorem: If  $S(X \times Y) > 0$ , then  $I(\xi, \eta) = \infty$ . If  $S(X \times Y) = 0$ , then

$$\begin{aligned} I(\xi, \eta) &= \int_X \int_Y a(x, y) \log a(x, y) P_{\xi}(dx) P_{\eta}(dy) \\ &= \int_X \int_Y \log a(x, y) P_{\xi\eta}(dx dy) \\ &= M \log \alpha_{\xi\eta}. \end{aligned} \quad (4)$$

Let us enumerate certain fundamental properties of  $I(\xi, \eta)$ .

- 1)  $I(\xi, \eta) = I(\eta, \xi)$ .
- 2)  $I(\xi, \eta) \geq 0$ ;  $I(\xi, \eta) = 0$  only if  $\xi$  and  $\eta$  are independent.
- 3) If the pair  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  are independent, then

$$I[(\xi_1, \xi_2), (\eta_1, \eta_2)] = I(\xi_1, \eta_1) + I(\xi_2, \eta_2).$$

- 4)  $I[(\xi, \eta), \zeta] \geq I(\xi, \zeta)$ .
- 5)  $I[(\xi, \eta), \zeta] = I(\eta, \zeta)$ , if and only if  $\xi, \eta, \zeta$  is a Markov sequence, i.e., if the conditional distribution of  $\zeta$  depends only on  $\eta$  for fixed  $\xi$  and  $\eta$ .

Apropos property 4, it is useful to note the following. In the case of the entropy

$$H(\xi) = I(\xi, \xi),$$

there is the bound on the entropy of the  $(\xi, \eta)$  pair from above:

$$H(\xi, \eta) \leq H(\xi) + H(\eta)$$

as well as the bound from below which results from 1 and 4.

$$H(\xi, \eta) \geq H(\xi); \quad H(\xi, \eta) \geq H(\eta).$$

A similar estimate for the quantity of information in  $\zeta$  relative to the  $(\xi, \eta)$  pair does not exist. From

$$I(\xi, \zeta) = 0; \quad I(\eta, \zeta) = 0$$

there still does not result the equality

$$I[(\xi, \eta), \zeta] = 0,$$

as can be shown by elementary examples.

For later use, let us note the special case when  $\xi$  and  $\eta$  are the random vectors:

$$\xi = (\xi_1, \dots, \xi_m)$$

$$\eta = (\eta_1, \dots, \eta_n) = (\xi_{m+1}, \dots, \xi_{m+n}),$$

and the quantities

$$\xi_1, \xi_2, \dots, \xi_{m+n}$$

are distributed normally with the second central moments

$$s_{ij} = M[(\xi_i - M\xi_i)(\xi_j - M\xi_j)].$$

<sup>3</sup> A. N. Kolmogorov, A. M. Iaglom, and I. M. Gel'fand, "Quantity of Information and Entropy for Continuous Distributions," Report at Third All-Union Math. Conf., 1956.

If the determinant

$$C = |s_{ij}|_{1 \leq i, j \leq m+n}$$

is not zero, then, as was calculated by Gelfand and Iaglom

$$I(\xi, \eta) = \frac{1}{2} \log \frac{AB}{C} \quad (5)$$

where

$$A = |s_{ij}|_{1 \leq i, j \leq m}; \quad B = |s_{ij}|_{m < i, j \leq m+n}.$$

It is often more expedient, however, to use another approach without the  $C > 0$  limitation. As is known,<sup>4</sup> all the second moments  $s_{ij}$  except those for which  $i = j$  or  $j = m + i$  go to zero after a suitable linear coordinate transformation in the  $X$  and  $Y$  spaces. For such a choice of coordinates

$$I(\xi, \eta) = -\frac{1}{2} \sum [1 - r^2(\xi_k, \eta_k)] \quad (6)$$

where the summation is taken over those

$$k \leq \min(m, n)$$

for which the denominator in the expression of the correlation coefficient

$$r(\xi_k, \eta_k) = \frac{s_{k, m+k}}{\sqrt{s_{kk} \cdot s_{m+k, m+k}}}$$

is not zero.

### III. ABSTRACT EXPLANATION OF THE PRINCIPLES OF THE SHANNON THEORY

Shannon considers the transmission of information according to the scheme

$$\xi \rightarrow \eta \rightarrow \eta' \rightarrow \xi'$$

where the "transmitting apparatus"

$$\eta \rightarrow \eta'$$

is characterized by the conditional distribution

$$P_{\eta'/\eta}(B'/y) = P(\eta' \in B' / \eta = y)$$

of the "output signal"  $\eta'$  for a given "input signal"  $\eta$  and a certain limitation

$$P_{\eta} \in V$$

of the input signal distribution  $P_{\eta}$ . The "coding"

$$\xi \rightarrow \eta$$

and "decoding" operations

$$\eta' \rightarrow \xi'$$

are characterized by the conditional distributions

$$P_{\eta/\xi}(B/x) = P(\eta \in B / \xi = x)$$

$$P_{\xi'/\eta'}(A'/y') = P(\xi' \in A' / \eta' = y').$$

The fundamental Shannon problem is the following. Given the spaces  $X, X', Y, Y'$  of possible values of the "input message"  $\xi$ , the "output message"  $\xi'$ , the input signal  $\eta$ , and the output signal  $\eta'$ ; given the characteristics of the transmitter, i.e., the conditional distribution  $P_{\eta'/\eta}$  and the class  $V$  of admissible input signal distributions  $P_{\eta}$ ; finally, given the distribution

$$P_{\xi}(A) = P(\xi \in A)$$

of the input message and the "fidelity criterion"

$$P_{\xi\xi'} \in W$$

where  $W$  is a certain class of joint distributions

$$P_{\xi\xi'}(C) = P[(\xi, \xi') \in C]$$

of the input and output communications. To find: Is it possible, and if it is, by what means, to give a coding and decoding rule (i.e., the conditional distributions  $P_{\eta/\xi}$  and  $P_{\xi'/\eta'}$ ) in such a manner that by calculating the distribution  $P_{\xi\xi'}$  in terms of the distributions  $P_{\xi}, P_{\eta/\xi}, P_{\eta'/\eta}, P_{\xi'/\eta'}$  under the assumption that the sequence

$$\xi, \eta, \xi', \eta'$$

is Markovian, we will obtain

$$P_{\xi\xi'} \in W?$$

As does Shannon, so let us define the "capacity" of the transmitter thus

$$C = \sup_{P_{\eta} \in V} I(\eta, \eta')$$

and let us introduce the quantity

$$H_w(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi')$$

which Shannon calls the "rate of creating information relative to a fidelity criterion" when computed per unit time. Then, the necessary condition of the possibility of transmission

$$H_w(\xi) \leq C \quad (7)$$

results at once from property 5 of Section II.

The incomparably deep idea of Shannon is that (7), when applied to the continuous operation of a "communication channel," is "almost sufficient" in a certain sense and under certain very broad conditions. From the mathematical point of view, it is a matter here of proving a limit theorem of the following type. It is assumed that the space  $X, X', Y, Y'$  of the distributions  $P_{\xi}$  and  $P_{\eta'/\eta}$  of the classes  $V$  and  $W$ , and therefore, of the quantities  $C$  and  $H_w(\xi)$ , depend on the parameter  $T$  (which plays the role in applications, of the duration of transmitter operation). It is required to establish that the condition

$$\liminf_{T \rightarrow \infty} \frac{C^T}{H_w^T(\xi)} > 1 \quad (8)$$

is sufficient, under a certain sufficiently general character of the assumptions, for the possibility of transmission satisfying the conditions formulated above, for sufficiently

<sup>4</sup> A. M. Obukhov, *Izv. AN USSR, Phys.-Math. Series*, pp. 339-370; 1938.

large  $T$ . Naturally, in such a formulation the problem is somewhat indistinct (for example, similar to the general problem of studying possible limit distributions for a sum of large numbers of "small" components). However, I intended to avoid any return to the terminology of the theory of stationary, random processes here, since it was shown in the note of the young Romanian mathematician Rozenblat-Rot Milu<sup>5</sup> that interesting results can be obtained in the designated direction without the assumption of stationariness.

Many remarkable works have been devoted to the derivation of a limit theorem of the kind indicated. The work of Khinchin<sup>6</sup> is the contribution of a USSR mathematician in this research direction. It appears to me that much remains to be done here. Namely, results of this kind are intended to give a foundation to the widespread conviction that the expression  $I(\xi, \eta)$  is not just one of the possible methods of measuring the "quantity of information" but it is a measure of the quantity of information having an advantage, in principle, over the others, actually. Since the "information," by its original nature, is not a scalar, then axiomatic investigations, permitting  $I(\xi, \eta)$  [or the entropy  $H(\xi)$ ] to be characterized uniquely by using simple formal properties, in this respect have lesser value, in my opinion. The situation here seems to me to be similar to our being ready to assign, at once, the greatest value to that method, out of all those proposed by Gauss to give a foundation to the normal law of error distribution, which starts from the limit theorem for the sum of a large number of small components. Other methods (for example, based on the principle of the arithmetic mean) demonstrate only why *any other* error distribution law could not be as acceptable and suitable as the normal law but they do not answer the question of why the normal law is actually encountered often in real problems. Similarly, the beautiful formal properties of the expressions  $H(\xi)$  and  $I(\xi, \eta)$  cannot demonstrate why they are sufficient for the complete (albeit from the asymptotic point of view) solution of many problems in many cases.

#### IV. CALCULATION AND ESTIMATION OF THE $\epsilon$ ENTROPY IN CERTAIN PARTICULAR CASES

If the condition

$$P_{\xi\xi', \epsilon} W$$

is chosen as the certainty of exact coincidence of  $\xi$  and  $\xi'$

$$P(\xi = \xi') = 1$$

then

$$H_w(\xi) = H(\xi).$$

In conformance with this, it seems to be natural to designate  $H_w(\xi)$  in the general case as the "entropy of the random object  $\xi$  for the accuracy of reproduction  $W$ ."

<sup>5</sup> Rozenblat-Rot Milu, *Trudy, Third All-Union Math. Conf.*, vol. 2, pp. 132-133; 1956.

<sup>6</sup> A. Ia. Khinchin, *Usp. Mate. Nauk*, vol. 11, no. 1 (67), pp. 17-75; 1956.

Now, let us assume that  $X$  is a metric space and that the space  $X'$  coincides with  $X$ , i.e., methods are investigated of the approximate transmission of information from the point  $\xi \in X$  by using the indication of the point  $\xi'$  of the same space  $X$ . It seems natural to require that

$$P\{\rho(\xi, \xi') \leq \epsilon\} = 1 \quad (W_\epsilon^0)$$

or that

$$M\rho^2(\xi, \xi') \leq \epsilon^2 \quad (W_\epsilon).$$

We will denote these two forms of the " $\epsilon$  entropy" of the distribution  $P_\xi$  by

$$H_{w_\epsilon^0}(\xi) = H_\epsilon^0(\xi)$$

$$H_{w_\epsilon}(\xi) = H_\epsilon(\xi).$$

As regards the  $\epsilon$  entropy  $H_\epsilon^0$ , I shall only note here a certain estimate for

$$H_\epsilon^0(X) = \sup_{P_\xi} H_\epsilon^0(\xi)$$

where the upper bound is taken over all the probability distributions  $P_\xi$  in the space  $X$ . As is known, for  $\epsilon = 0$ ,

$$H_0^0(X) = \sup_{P_\xi} H(\xi) = \log N_x$$

where  $N_x$  is the number of elements of the manifold  $X$ . For  $\epsilon > 0$ ,

$$\log N_x^c(2\epsilon) \leq H_\epsilon^0(x) \leq \log N_x^a(\epsilon)$$

where  $N_x^a(\epsilon)$  and  $N_x^c(\epsilon)$  are characteristics of the space  $X$  which are introduced in my note.<sup>2</sup> The asymptotic properties of the function  $N_x(\epsilon)$  as  $\epsilon \rightarrow 0$ , studied in my work<sup>2</sup> for a number of specific spaces  $X$ , are interesting analogs of the properties, explained later, of the asymptotic behavior of the function  $H_\epsilon(\xi)$ .

Let us now turn to the  $\epsilon$  entropy  $H_\epsilon(\xi)$ . If  $X$  is an  $n$ -dimensional Euclidean space and if

$$P_\xi(A) = \int_A P_\xi(x) dx_1 dx_2 \cdots dx_n$$

then, at least in the case of the sufficiently smooth function  $p_\xi(x)$ , the following well-known formula holds:

$$H_\epsilon(\xi) = n \log \frac{1}{\epsilon} + [h(\xi) - n \log \sqrt{2\pi e}] + o(1) \quad (9)$$

where

$$h(\xi) = - \int_X p_\xi(x) \log p_\xi(x) dx_1 \cdots dx_n$$

is the "differential entropy," already introduced in the first Shannon works. Hence, the asymptotic behavior of  $H_\epsilon(\xi)$  in the case of sufficiently smooth continuous distributions in  $n$ -dimensional space is determined, to a first approximation, by the dimensionality of the space and the differential entropy  $h(\xi)$  only enters as the second term in the expression for  $H_\epsilon(\xi)$ .

It is natural to expect that the growth of  $H_\epsilon(\xi)$  as  $\epsilon \rightarrow 0$  will be substantially more rapid for typical distributions

in infinite-dimensional spaces. As the simplest example, let us consider the Wiener random function  $\xi(t)$ , defined for  $0 \leq t \leq 1$ , with the normally-distributed independent increments:

$$\Delta\xi = \xi(t + \Delta t) - \xi(t)$$

for which

$$\xi(0) = 0; \quad M\Delta\xi = 0; \quad M(\Delta\xi)^2 = \Delta t.$$

Iaglom found that in this case, in the  $L^2$  metric

$$H_\epsilon(\xi) = \frac{4}{\pi} \frac{1}{\epsilon^2} + o\left(\frac{1}{\epsilon^2}\right) \quad (10)$$

Under certain natural assumptions, the formula

$$H_\epsilon(\xi) = \frac{4}{\pi} \chi\left(\frac{1}{\epsilon}\right) + o\left(\frac{1}{\epsilon}\right) \quad (11)$$

where

$$\chi(\xi) = \int_{t_0}^{t_1} Mb | t, \xi(t) | dt$$

can be obtained in a more general way for the diffuse kind of Markov process on the  $t_0 \leq t \leq t_1$  time segment with

$$M_{\Delta\xi} = A[t, \xi(t)] \Delta t + o(\Delta t); \\ M(\Delta\xi)^2 = B[t, \xi(t)] \Delta t + o(\Delta t).$$

The  $\epsilon$ -entropy  $H_\epsilon$  can be calculated exactly for the case of the normal distribution in an  $n$ -dimensional space or in Hilbert space. After a suitable orthogonal coordinate transformation, the  $n$ -dimensional vector  $\xi$  assumes the form

$$\xi = (\xi_1, \xi_2, \dots, \xi_n)$$

where the coordinates  $\xi_k$  are mutually independent and distributed normally. The parameter  $\theta$ , for given  $\epsilon$ , is determined from the equation

$$\epsilon^2 = \sum \min(\theta^2, D^2\xi_k)$$

and in the case of  $\xi$  distributed normally

$$H_\epsilon(\xi) = \frac{1}{2} \sum_{D^2\xi_k \rightarrow \theta^2} \log \frac{D^2\xi_k}{\theta^2}. \quad (12)$$

The approximating vector

$$\xi' = (\xi'_1, \xi'_2, \dots, \xi'_n)$$

should be chosen such that

$$\xi'_k = 0$$

for  $D^2\xi_k \leq \theta^2$  and

$$\xi_k = \xi'_k + \Delta_k; \quad D^2\Delta_k = \theta^2; \quad D^2\xi'_k = D^2\xi_k - \theta^2$$

for  $D^2\xi_k > \theta^2$  and the vectors  $\xi_k$  and  $\Delta_k$  are mutually independent. The infinite-dimensional case is in no way different from the finite dimensional.

Finally, it is very essential that the *maximum value of  $H_\epsilon(\xi)$  for the vector  $\xi$  ( $n$  dimensional or infinite dimensional)*

*be attained in the normal distribution case for given second central moments.* This result can be obtained directly or from the following proposition of Pinsker.<sup>7</sup>

Theorem: Let the positive-definite symmetric matrix of the  $s_{ij}$ ,  $0 \leq i, j \leq m+n$  quantities and the distribution  $P_\xi$  of the vector be given

$$\xi = (\xi_1, \xi_2, \dots, \xi_m)$$

for which the central second moments equal  $s_{i,j}$  (for  $0 \leq i, j \leq m$ ). Let the condition  $W$  on the joint distribution  $P_{\xi\xi'}$  of the vector  $\xi$  and the vector

$$\xi' = (\xi_{m+1}, \xi_{m+2}, \dots, \xi_{m+n})$$

be that the central second moments of the quantities

$$\xi_1, \xi_2, \dots, \xi_{m+n}$$

equal  $s_{ij}$  (for  $0 \leq i, j \leq m+n$ ). Then

$$H_W(\xi) \leq \frac{1}{2} \log \frac{AB}{C}. \quad (13)$$

The notation in (13) corresponds to the explanation of Section II. It is seen from a comparison with the results of Section II, that inequality (13) becomes the equality in the case of the normal distribution  $P_\xi$ .

The principles of solving the variational problems arising in the calculation of the "rate of creating information" were indicated sufficiently long ago by Shannon. Shannon and Weaver<sup>8</sup> write: "Unfortunately these formal solutions are difficult to evaluate in particular cases and seem to be of little value."<sup>9</sup> In substance, however, many problems of this kind are simple enough, as is seen from the above. It is possible that the slow development of investigations in this direction is related to insufficient understanding of the fact that the solution of the variation problem often appears to be degenerate in typical cases: For example, the evaluation of  $H_\epsilon(\xi)$  in the problem selected above, for the normally distributed vector  $\xi$  in the  $n$ -dimensional case, the vector  $\xi'$  often appears to be not  $n$  dimensional but only  $k$  dimensional with  $k < n$ ; in the infinite-dimensional case, the vector  $\xi'$  always appears to be finite dimensional.

## V. QUANTITY OF INFORMATION AND RATE OF CREATING INFORMATION IN THE STATIONARY PROCESS CASE

Let us consider two stationary and stationarily-related processes,

$$\xi(t), \eta(t) \quad -\infty < t < +\infty.$$

Let us denote by  $\xi_T$  and  $\eta_T$  the segments of the  $\xi$  and  $\eta$  processes in the time  $0 < t \leq T$  and by  $\xi_-$  and  $\eta_-$  the flow of the  $\xi$  and  $\eta$  processes on the negative semiaxis  $-\infty < t \leq 0$ . To give the pair  $(\xi, \eta)$  of the stationarily-related  $\xi$  and  $\eta$  processes means to give the probability distribution

<sup>7</sup> M. S. Pinsker, *Trudy, Third All-Union Math. Conf.*, vol. 1, p. 125; 1956.

<sup>8</sup> C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press; 1949.

<sup>9</sup> *Ibid.*, sec. 28, p. 79 of the Russian translation.

$P_{\xi\eta}$ , invariant to shift along the  $t$  axis, in the space of the function pair  $\{x(t), y(t)\}$ . If  $\xi_-$  is fixed, then the following conditional probability

$$P_{\xi_T\eta/\xi_-}(C/x_-) = P\{(\xi_T, \eta) \in C/\xi_- = x_-\}$$

arises from the distribution  $P_{\xi\eta}$ . Using this distribution, the conditional quantity of information

$$I(\xi_T, \eta/x_-)$$

is calculated in conformance with Section II. If the mathematical expectation

$$MI(\xi_T, \eta/\xi_-)$$

is finite for any  $T > 0$ , then it is finite for all other  $T > 0$  and

$$MI(\xi_T, \eta/\xi_-) = T\bar{I}(\xi, \eta).$$

It is natural to call the quantity  $\bar{I}(\xi, \eta)$  the "rate of creating information of the process  $\eta$  for compliance with the process  $\xi$ ." If the process  $\xi$  can be extrapolated with complete accuracy to future occurrences, then

$$\bar{I}(\xi, \eta) = 0.$$

In particular, this will be so if the process  $\xi$  has a bounded spectrum. Generally speaking, the following equality

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi) \quad (14)$$

does not hold. However, under sufficiently broad conditions on the "regularity" of the process  $\xi$ ,<sup>10</sup> the equality

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi)$$

holds, where

$$\bar{I}(\xi, \eta) = \lim_{T \rightarrow \infty} \frac{1}{T} I(\xi_T, \eta_T).$$

Since  $I(\xi_T, \eta_T) = I(\eta_T, \xi_T)$ , then always

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi)$$

and, therefore, when both equalities  $\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi)$  and  $\bar{I}(\eta, \xi) = \bar{I}(\eta, \xi)$  are correct, equality (14) holds. Now, let  $W$  be a certain class of joint distributions  $P_{\xi\xi'}$  of two stationary and stationarily-related processes  $\xi$  and  $\xi'$ . It is natural to call the quantity

$$\bar{H}_W(\xi) = \inf_{P_{\xi\xi'} \in W} \bar{I}(\xi', \xi)$$

the "rate of creating information in the process  $\xi$  under the accuracy of reproduction  $W$ ." It can be shown that

$$\bar{H}_W(\xi) = \bar{H}_W(\xi)$$

where

$$\bar{H}_W = \inf_{P_{\xi\xi'} \in W} \bar{I}(\xi, \xi')$$

<sup>10</sup> Here and later, the regularity of the process means, roughly speaking, that the segments of the process, corresponding to two segments of the  $t$  axis sufficiently removed from each other, are almost independent. In the case of Gaussian processes, the well-known definition of regularity introduced in my work<sup>14</sup> is applicable here.

under certain assumptions on the regularity of the process  $\xi$  and for certain natural types of conditions  $W$ .

## VI. CALCULATION AND ESTIMATION OF THE AMOUNT OF INFORMATION AND THE RATE OF CREATING INFORMATION IN TERMS OF THE SPECTRUM

Pinsker<sup>11</sup> established the formula:

$$\bar{I}(\xi, \eta) = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \log [1 - r^2(\lambda)] d\lambda \quad (15)$$

where

$$r^2(\lambda) = \frac{|f_{\xi\eta}(\lambda)|^2}{f_{\xi\xi}(\lambda)f_{\eta\eta}(\lambda)}$$

and  $f_{\xi\xi}$ ,  $f_{\xi\eta}$ ,  $f_{\eta\eta}$  are spectral densities; for the case when the distribution  $P_{\xi\eta}$  is normal and at least one of the processes  $\xi$  or  $\eta$  is regular. In connection with the review by Doob,<sup>12</sup> we would like to note that the novelty, in principle, of the Pinsker result is somewhat greater than can be expected on the basis of this review. The expression

$$\bar{h}(\xi) = \log (2\pi\sqrt{e}) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_{\xi\xi}(\lambda) d\lambda \quad (16)$$

is known in the case of processes with discrete time  $t$  for the differential entropy of a normal process per unit time:

$$\bar{h}(\xi) = \lim_{T \rightarrow \infty} \frac{1}{T} h(\xi_1, \xi_2, \dots, \xi_T).$$

However, no analog of the expression  $\bar{h}(\xi)$  exists in the continuous time and unbounded spectrum case and the Pinsker formula requires independent derivation.

It is natural to characterize the accuracy of reproducing the stationary process  $\xi$ , using the stationary process  $\xi'$  stationarily related to  $\xi$ , by the quantity

$$\sigma^2 = M[\xi(\alpha) - \xi'(\alpha)]^2$$

and in the case of a  $W$  condition of the form

$$\sigma^2 \leq \epsilon^2$$

it is natural to call the quantity

$$\bar{H}_\epsilon(\xi) = \bar{H}_W(\xi)$$

the  $\epsilon$  entropy per unit time of the process  $\xi$  and under the assumption that

$$\bar{H}_W(\xi) = \bar{H}_W(\xi)$$

the rate of creating information in the process  $\xi$  for average accuracy of transmission  $\epsilon$ . It can be concluded from the appropriate statement for finite-dimensional distributions (see Section IV) that the quantity  $\bar{H}_\epsilon(\xi)$  attains a maximum in the case of the normal process  $\xi$  for a given spectral density  $f_{\xi\xi}(\lambda)$ . In the normal case,  $\bar{H}_\epsilon(\xi)$  can be calculated easily in terms of the spectral density exactly as was explained in Section IV applied to  $H_\epsilon(\xi)$  for the  $n$ -di-

<sup>11</sup> M. S. Pinsker, *Doklady, AN USSR*, vol. 98, 213-216; 1954.

<sup>12</sup> J. L. Doob, *Math. Revs.*, vol. 16, p. 495; 1955.

mensional distribution. The parameter  $\theta$  is determined from the equation

$$\epsilon^2 = \int_{-\infty}^{\infty} \min [\theta^2, f_{\xi\xi}(\lambda)] d\lambda. \quad (17)$$

Using this parameter, the quantity  $\bar{H}_\epsilon(\xi)$  is found from the formula

$$H_\epsilon(\xi) = \frac{1}{2} \int_{f_{\xi\xi}(\lambda) \theta^2} \log \frac{f(\lambda)}{\theta^2} d\lambda. \quad (18)$$

Spectral densities of the kind shown in Fig. 1 which are approximated well by the function:

$$\varphi(\lambda) = \begin{cases} a^2 & \text{for } A \leq |\lambda| \leq A+W \\ 0 & \text{in the remaining cases} \end{cases}$$

are of practical interest. It is easy to calculate that

$$\theta^2 \sim \frac{\epsilon^2}{2W} \quad (19)$$

$$\bar{H}_\epsilon(\xi) \sim W \log \frac{2Wa^2}{\epsilon^2}$$

approximately, in this case for *not too small*  $\epsilon$  for a normal process.

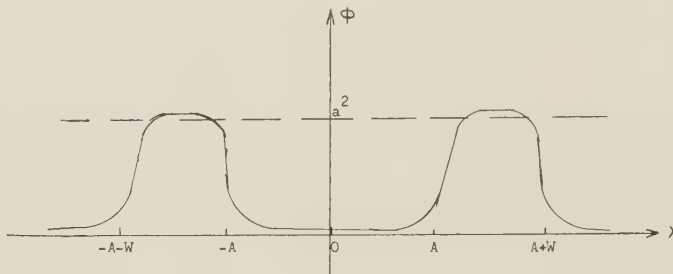


Fig. 1

Certainly, (19) is none other than the well-known Shannon formula

$$R = W \log \frac{Q}{N}. \quad (20)$$

Here, however, the novelty, in principle, is that now we see why and within what limits (for not too small  $\epsilon$ ) this formula can be used for a process with an unbounded spectrum and such are all the processes in the theory of information transmission which really interest us.

Writing (19) thus

$$\bar{H}_\epsilon(\xi) \sim 2W \log (a \sqrt{2W}) \log \frac{1}{\epsilon} \quad (21)$$

and comparing with (9), we see that the double width  $2W$  of the useful frequency band plays the role of the number of measurements. This idea of the equivalence of twice the frequency bandwidth to the number of measurements occurring in a certain sense of the word, per unit time was apparently first expressed by Kotel'nikov.<sup>13</sup> On the basis of this idea, Kotel'nikov indicated the fact that a function, for which the spectrum is limited to bandwidth  $2W$ , is determined uniquely by the values of the function at the points

$$\dots, -\frac{2}{2W}, -\frac{1}{2W}, 0, \frac{1}{2W}, \frac{2}{2W}, \dots, \frac{k}{2W}, \dots$$

Shannon retained this argumentation, using the representation obtained in this manner to derive (20). Since a function with a bounded spectrum is always singular in the sense of my work<sup>14</sup> and the observation of such a function is not related, generally, to the stationary flow of new information, then the sense of this kind of argumentation does not remain completely clear so that the new derivation of the *approximate* formula (21), cited here, seems to me to be not devoid of interest.

The growth of  $\bar{H}_\epsilon(\xi)$  as  $\epsilon$  decreases occurs, for small  $\epsilon$ , for any normally distributed regular random function substantially more rapidly than would be obtained according to (21). In particular, if  $f_{\xi\xi}(\lambda)$  has order  $1/\lambda\beta$  as  $\lambda \rightarrow \infty$ , then  $\bar{h}_\epsilon(\xi)$  has order  $1/(2/\epsilon(\beta - 1))$ .

<sup>13</sup> V. A. Kotel'nikov, Material for the First All-Union Conf. on questions of communications; 1933.

<sup>14</sup> A. N. Kolmogorov, Bulletin, Moscow Univ. I, no. 6; 1941. English translation available.



# On Noise Stability of a System with Error-Correcting Codes\*

VLADIMIR I. SIFOROV†

**Summary**—The problem posed in this paper is to give a relation connecting the noise stability of a communication system in which error-correcting codes are used to the parameters of these codes. The amount  $x$  of code combinations is found, which differ from each other by not less than a given number of elements for a given arrangement of all the possible combinations. It is proved that the quantity  $x$  depends on the arrangement of the primary combinations. Inequalities are obtained for the greatest amount of combinations  $x_m$ , for which any two differ by not less than  $d$  elements for small and large values  $n$  of the common number of elements in each code combination. It is established that the probability of distortion,  $p_n$ , of a code group in a system with correcting codes satisfies the inequality  $p_n < f(p, n, d)$ , where  $p$  is the probability of distortion of one element. The shape of the  $f(p, n, d)$  function is found for large values of  $n$ . Graphs of this function are constructed.

## INTRODUCTION

ONE OF THE means of guaranteeing the high noise-stability and capacity of a communication system is to use error-correcting codes. The usual principle for constructing these codes is that additional symbols, intended to disclose and to correct the errors caused by noise, are added to code combinations of the usual kind.

Shannon<sup>1</sup> showed that if any  $2^{C_n}$  combinations are chosen at random from all  $2^n$  possible combinations of  $n$ -elementary dyadic signals, then the frequency of the error, caused by the noise, can be made as small as desired for  $C < C_0$  and sufficiently large values of  $n$ . In order to guarantee the greatest system capacity for the  $n$  value selected and for the sufficiently small error frequency given, the choice of the coding combinations must be made in such a way that they are as different as possible from each other.

Correcting codes based on the use of only parts of combinations of elementary dyadic signals from all possible combinations, were investigated in a number of works, among which, for example, are those of Hamming,<sup>2</sup> Laemmel,<sup>3</sup> Reed,<sup>4</sup> and Silverman and Basler.<sup>5</sup>

\* Presented at 1956 Symposium on Information Theory at Mass. Inst. Tech., Cambridge, Mass., September 10-12, 1956.

† Corresponding member, Academy of Science, USSR.

<sup>1</sup> C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, p. 349; 1948.

<sup>2</sup> R. W. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147-160; 1950.

<sup>3</sup> A. E. Laemmel, "Efficiency of Noise-Reducing Codes" and "Communication Theory," papers read at a symposium on "Applications of Communication Theory" held at IEE, London, September 22-26, 1952; also *Butterworths Sci. Publ.*, pp. 111-118; 1953.

<sup>4</sup> Reed, "A Class of Multiple-Error-Correcting Codes and the Detecting Scheme, M.I.T. Lincoln Lab. Tech. Rep. no. 44; 1953.

<sup>5</sup> R. A. Silverman and M. Basler, "Coding for constant-data-rate systems-part I, a new error-correcting code," *Proc. IRE*, vol. 42, pp. 1428-1435; September, 1954, and "part II—multiple-error-correcting codes," *Proc. IRE*, vol. 43, pp. 728-733; June, 1955.

However, relations connecting the noise stability of the transmission system, in which the correcting codes are used, to the parameters of these codes, are not found in general-enough form in these and other works known to the author. The search for such general relations is a very difficult problem which must be solved by investigating a number of questions from the region of multi-dimensional geometry and number theory. An attempt is made in this paper to give a partial solution of the problem mentioned.

## ON THE NUMBER OF CODE COMBINATIONS DIFFERING FROM EACH OTHER BY NOT LESS THAN A GIVEN NUMBER OF ELEMENTS (SMALL $n$ CASE)

Let us consider the code combinations of  $n$  elements, each of which has the value 0 or 1. The common number of all possible combinations of this kind will be  $2^n$ . From these, let us select only those which satisfy the condition that any two of the selected combinations will differ from each other by not less than  $d$  elements. Let us call the quantity  $d$  the distance between the combinations. Let us denote by  $x_m$  the greatest possible number of combinations selected which satisfy the condition mentioned above.

Therefore,  $x_m$  is the greatest possible number of combinations for which any two are separated by a distance equal to or greater than  $d$ . Evidently,  $x_m$  is a function of  $n$  and  $d$ , i.e.,

$$x_m = f(n, d). \quad (1)$$

Let us explain the definition cited here by a specific example. Let  $n = 3$  and  $d = 2$ . The set of all possible combinations will be: 000, 100, 010, 110, 001, 101, 011, 111. The distances between any two of these eight combinations are within the limits from 1 to 3. Discarding the combinations separated by the unit distance, we easily confirm that there are only four combinations, satisfying the condition presented, i.e.,

$$x_m = f(3, 2) = 4.$$

For example, these combinations are: 000, 110, 101, and 011. Any two of these differ from each other by not less than two elements.

Any two of the selected set of combinations differ from each other by not less than three elements for  $d = 3$ ,  $n \geq 3$ . When such a set is used to transmit information, the distortion caused by noise can be corrected automatically in any one element. Actually, the distorted combination will differ for  $d = 3$  from the true by one

element and from any other, by two or more elements. In other words, the distorted combination will be closer to the real for a single error than to the false. In principle, this affords the possibility of correcting single errors automatically.

Similarly, it appears to be possible, for  $d = 5$ , to correct the errors in two elements automatically, for  $d = 7$ , in three and, in general, for any odd  $d$  in  $(d - 1)/2$  elements. As is easily verified, automatic correction of  $(d/2 - 1)$  elements is possible for even values of  $d$ . For example, for  $d = 4$ , correction is possible of only a single error and for  $d = 8$ , correction of errors in three elements.

The problem of searching for the  $x_m = f(n, d)$  function can easily be interpreted geometrically as the problem of finding the greatest number of vertices of an  $n$ -dimensional "cube" for which any two differ from each other by not less than  $d$  coordinates.

Shown in Fig. 1 is a three-dimensional cube ( $n = 3$ ) and also the four vertices found above (000, 110, 101, and 011) for which any two differ from each other by not less than two coordinates ( $d = 2$ ).

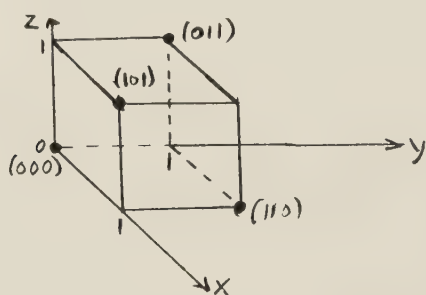


Fig. 1

The number of code combinations which differ from each other in not less than given number of elements can be found for small values of  $n$  and  $d$  by direct computation.

Let us assume that all  $2^n$  possible combinations are arranged in a series and that each has an index number. Let us compare each of the combinations, starting with the second and ending with the  $2^n$ -th, with the first. Let us cancel those which differ from the first in less than  $d$  elements. Now, let us enumerate the remaining combinations. Evidently, the greatest index number here will be less than  $2^n$ . Now, let us compare each of the remaining combinations, starting with the third and ending with the last, with the second. Let us cancel those which differ from the second in less than  $d$  elements. Let us enumerate the remaining combinations and let us compare each of them, starting with the fourth and ending with the last, with the third. Let us continue this procedure further in a similar manner and let us end it by comparing the last of the remaining combinations with the preceding.

Let us denote by  $x$  the number of all combinations obtained as a result of the procedure described for given values of  $n$  and  $d$ . Evidently

$$x = F(n, d). \quad (2)$$

In general  $x$  depends not only on  $n$  and  $d$  but on the initially-selected arrangement of all  $2^n$  primary combinations. The  $x_m$  quantity which we introduced before in (1) is the greatest value of the set of  $x$  values corresponding to all possible permutations of the  $2^n$  primary combinations for which the number is  $2^n!$ .

Let us find  $x$  for a definite arrangement of the primary combinations in correspondence with the following procedure. Let us write all the possible combinations for  $n = 1$ . These combinations are 0 and 1. Let us add to each of them first 0 and then 1. Then, we obtain a set of all possible combinations for  $n = 2$  as: 00, 10, 01, and 11. Again let us add first 0 and then 1 to each of these combinations. Hence, we obtain the set of all possible combinations for  $n = 3$  as

$$000, 100, 010, 110, 001, 101, 011, 111.$$

Adding first 0 and then 1 to each of these combinations, we obtain the set of all possible combinations for  $n = 4$ . Continuing this procedure further, we find the set of all primary combinations for  $n = 5, 6, 7$ , etc.

E. N. Zotova determined  $x$  in (2) for the arrangement described here of the primary combinations and calculated the set of selected combinations of interest by the above-mentioned procedure. The results of these computations are given in Table I.

TABLE I  
VALUES OF THE  $x = F(n, d)$  FUNCTION

$n \backslash d$	1	2	3	4	5	6	7	8	9	10
1	2	4	8	16	32	64	128	256	512	1024
2	—	2	4	8	16	32	64	128	256	512
3	—	—	2	4	8	16	32	64	128	256
4	—	—	—	2	4	8	16	32	64	128
5	—	—	—	—	2	4	8	16	32	64
6	—	—	—	—	—	2	4	8	16	32
7	—	—	—	—	—	—	2	4	8	16
8	—	—	—	—	—	—	—	2	4	8
9	—	—	—	—	—	—	—	—	2	4
10	—	—	—	—	—	—	—	—	—	2

In order to determine the degree of influence of the arrangement of the primary combinations on  $x$ , Zotova calculated this quantity for four pairs of fixed values of  $n$  and  $d$  and for ten different arrangements of the primary combinations. The results of these calculations are shown in Table II.

TABLE II  
VALUES OF  $x = F(n, d)$  FOR VARIOUS ARRANGEMENTS OF THE PRIMARY COMBINATIONS

Test No.	1	2	3	4	5	6	7	8	9	10
Values of $n, d$										
$n = 8; d = 3$	16	16	17	17	17	16	15	19	16	16
$n = 8; d = 4$	16	11	10	10	8	8	7	9	8	9
$n = 6; d = 3$	8	8	6	6	6	8	6	6	6	6
$n = 6; d = 4$	4	2	2	4	2	4	4	2	2	4

The data of these tables confirm the hypothesis cited above that the arrangement of the primary combinations affects the number of sets of combinations, sorted by the procedure described, for which any two differ from each other in not less than a given number of elements.

It is easy to confirm that  $x$  depends on the arrangement of the primary combinations even for small values of  $n$  and  $d$  and, in particular, for  $n = 3$  and  $d = 2$ . Actually, let us write the primary combinations for this case as

$$000, 111, 100, 010, 110, 001, 101, 011.$$

Using the cancellation procedure described above, we find that only two combinations remain, 000 and 111, which satisfied the condition presented. In other words, for the given arrangement of primary combinations  $x(3, 2) = 2$  while it was shown above that for the arrangement established before  $x(3, 2) = 4$ .

The geometric interpretation of this statement is given in Fig. 2. Here, two points are shown which correspond

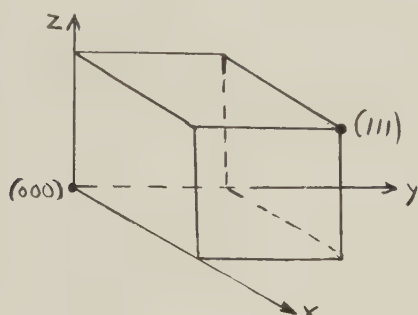


Fig. 2—Geometric interpretation of the relation  $x(3, 2) = 2$ .

to the two primary combinations 000 and 111 of sequence of eight primary combinations. The addition of any of the remaining six points to these two will, evidently, lead to the appearance of a pair of points different from each other not in two, but only in one coordinate.

From Tables I and II it is easy to find the lower limit of  $x_m$  in which we are interested, for small  $n$  and  $d$  values. Thus, for example, on the basis of the data of Table II it can be stated that

$$x(8, 3) \geq 19.$$

Laemmel, on the basis of the investigations of Hamming<sup>2</sup> and of other authors, gives relations which when used enables the limits between which the values of the  $x_m(n, d)$  function lies to be determined. These relations are the following:

$$\frac{2^n}{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{2\Delta}} \leq x_m(n, 2\Delta + 1) \leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{\Delta}} \quad (3)$$

$$x_m(n, 2\Delta) = x_m(n - 1, 2\Delta - 1) \quad (4)$$

$$\frac{1}{2} x_m(n + 1, d) \leq x_m(n, d) \leq x_m(n + 1, d) \quad (5)$$

$$x_m(n, d + 1) \leq x_m(n, d) \quad (6)$$

$$x_m(n_1, d_1), x_m(n_2, d_2) \leq x_m(an_1 + bn_2, ad_1 + bd_2) \quad (7)$$

$$x_m(n, 1) = 2^n; \quad x_m(n, 2) = 2^{n-1} \quad (8)$$

$$2^{n-\gamma} \leq x_m(n, 3) \quad (9)$$

$$x_m(n, d) = 2 \quad \text{for } d > \frac{2}{3}n \quad (10)$$

$$x_m(n, d) = 4 \quad \text{for } d = \frac{2}{3}n. \quad (11)$$

Here  $\Delta$  is the largest number of elements in which the distortion can be corrected automatically in each code combination;

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots$$

are the binomial coefficients

$$1, \frac{n}{1}, \frac{n(n-1)}{1 \cdot 2},$$

etc., respectively;  $a$  and  $b$  are any positive integers;  $\gamma$  is the least integer satisfying the inequality

$$\gamma \geq \log_2(n + 1). \quad (12)$$

The expression on the right side of the double inequality (3) was obtained by Hamming<sup>2</sup> as the quotient of dividing the common number of  $2^n$  points lying on a sphere in  $n$ -dimensional space by the number of points contained in a small sphere representing one code combination.

Setting  $n = 8$  and  $\Delta = 1$  in (3), we obtain

$$6 \leq x_m(8, 3) \leq 28.$$

From Tables I and II above, it was found that  $x_m(8, 3) \geq 19$ . Consequently, it can be said that

$$19 \leq x_m(8, 3) \leq 28.$$

Using relations (3) to (11) and Tables I and II, the upper and lower limits between which the desired function  $x_m(n, d)$  lies for small  $n$  and  $d$  values can be determined in an analogous way.

#### NOISE STABILITY OF A SYSTEM WITH ERROR-CORRECTING CODES FOR A LARGE NUMBER OF ELEMENTS

The formulas and tables cited in the preceding section do not permit a representation of the character of the  $x_m(n, d)$  function to be formed in the region of large  $n$  values. Consequently, in this case they do not afford the possibility of calculating the capacity and the noise stability of a system with correcting codes.

Let us formulate the problem of finding the inequality which can be of use in assessing the noise stability of a system with correcting codes for the case when the number  $n$  of elements is large.

Let us denote the probability of the distortion, caused by noise, of any one element through  $p$ . In other words,  $p$  is the probability of zero transforming into one or one into zero in one element of the transmitted code combination.

When a sequence of positive and negative currents (Fig. 3) are transmitted in the presence of noise, then  $p$  is the probability of the negative current transforming into a positive or a positive into a negative.

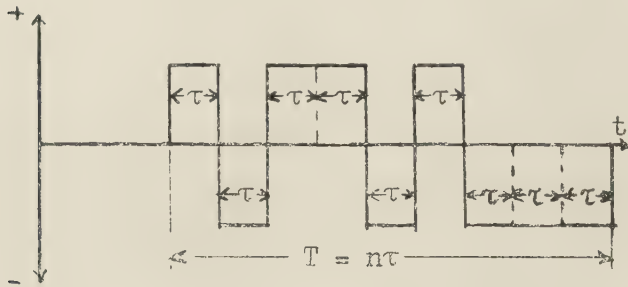


Fig. 3

Let us assume that the noise acts independently on each element of communication. In other words, we will assume that the probability of distortion of a given element by noise is not completely dependent on how all the remaining elements are subjected to distortion.

Let us take a set of code combinations, each of which consists of  $n$  elements. Let us assume that any two combinations of this set differ from each other by not less than in  $d$  elements.

When such a set is used to transmit information under noise conditions, each combination, generally, will be distorted by the noise. If any combination will be distorted to a small degree, then it will be transmitted without distortion if automatic error correction is used. If the degree of distortion of the transmitted combination is large, then the distorted combination can appear to be closer to the false than to the true. In this case, an error in the information transmission will occur even if automatic correction is used. Let us denote the probability of such an error by  $p_n$  and let us pose the problem of finding the relation between this probability  $p_n$  and the probability  $p$  of the distortion of a single element of information, the number  $n$  of all elements and the code distance  $d$ .

Let us denote by  $\mu$  the number of distorted elements in one combination composed of  $n$  elements. If

$$\mu < \frac{d}{2} \quad (13)$$

then, evidently, the distorted combination will be closer to the true than to any other and the information transmission will be without error. If

$$\mu \geq \frac{d}{2} \quad (14)$$

then the distorted combination can appear to be closer either to the true or to the false depending on the character of the distortion and the properties of the set of code combinations selected.

Actually, let us take the set of combinations shown in Table III, for example.

TABLE III  
CODE COMBINATIONS FOR  $n = 6$  AND  $d = 3$

Index number	1	2	3	4	5	6
Combination	100001	001110	111000	010101	011011	110100

Let us assume that combination no. 3 is transmitted and that  $\mu = 2$ . Let us admit that the fourth and fifth elements of this combination are subjected to distortion. Then the distorted combination will be 111110. Comparing it with all six combinations shown in Table III we obtain the respective differences in the numbers of the elements: 5, 2, 2, 4, 3, and 1. Hence, it is seen that the distorted combination no. 3 appears to be closest to combination no. 6. Consequently, the information transmission in the given case will be with error.

Now, let the first and second elements in the transmitted combination no. 3 be subjected to distortion. Then the distorted combination will be 001000. Comparing it again with all six combinations of the table, we obtain the respective differences in the number of elements: 3, 2, 2, 4, 3, and 5. Hence, it is seen that the distorted combination no. 3 appears to be closest to combination no. 2 and no. 3. Consequently, here the information transmission will be either with error or without error depending on whether the distorted combination no. 3 is transformed into combination no. 2 or no. 3.

Let us take a code combination of  $n$  elements. In conformance with the Moivre-Laplace integral theorem, the number of elements distorted by noise satisfies the relation

$$P_{n \rightarrow \infty} \left\{ a_1 \leq \frac{\mu - np}{\sqrt{np(1-p)}} < a_2 \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{a_1}^{a_2} \exp \left[ -\frac{z^2}{2} \right] dz \quad (15)$$

where  $P$  is the probability of conserving the inequality in the braces.

Here, putting  $a_1 = -\infty$ ,  $a_2 = a$  and taking into account that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^a \exp \left[ -\frac{z^2}{2} \right] dz = 1 - \frac{1}{\sqrt{2\pi}} \int_a^{\infty} \exp \left[ -\frac{z^2}{2} \right] dz$$

we obtain

$$P_{n \rightarrow \infty} \left\{ \frac{\mu - np}{\sqrt{np(1-p)}} < a \right\} \rightarrow 1 - \frac{1}{\sqrt{2\pi}} \int_a^{\infty} \exp \left[ -\frac{z^2}{2} \right] dz. \quad (16)$$

TABLE IV  
CODE DISTANCE  $d$  FOR VARIOUS VALUES OF  $p_n$  AND  $p$

	$p$	0.01	0.02	0.05	0.1	0.2	0.3	0.4	0.5
	$n$								
$p_n = 0.1$	10	1.006	1.534	2.766	4.432	7.242	9.714	11.97	14.05
	100	4.550	7.588	15.58	27.69	50.25	71.74	92.75	112.8
	1000	28.06	51.35	117.7	224.3	432.4	637.1	839.7	1041
	10000	255.5	435.9	1056	2077	4102	6117	8126	10130
$p_n = 0.01$	10	1.663	2.459	4.206	6.414	9.885	12.74	15.21	17.36
	100	6.628	10.51	20.14	33.96	58.61	81.32	102.8	123.3
	1000	34.63	60.59	132.1	244.1	458.8	667.4	872.0	1074
	10000	246.3	465.1	1101	2140	4186	6213	8228	10230
$p_n = 0.001$	10	2.144	3.136	5.258	7.863	11.82	14.95	17.57	19.7
	100	8.148	12.65	23.47	38.54	64.72	88.32	110.3	130.9
	1000	39.44	67.36	142.6	258.6	478.2	689.5	895.7	1098
	10000	261.5	486.5	1135	2185	4247	6283	8303	10310
$p_n = 0.0001$	10	2.540	3.692	6.125	9.056	13.41	16.78	19.52	21.76
	100	9.400	14.41	26.21	42.31	69.75	94.08	116.6	137.2
	1000	43.40	72.92	151.2	270.6	494.1	717.8	915.2	1118
	10000	274.0	504.1	1162	2223	4298	6341	8364	10370

Let us select  $a$  in conformance with the equality

$$a = \frac{d/2 - np}{\sqrt{np(1-p)}}. \quad (17)$$

Then (16) can be represented thus:

$$P_{n \rightarrow \infty} \left\{ \mu < \frac{d}{2} \right\} \rightarrow 1 - \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp \left[ -\frac{z^2}{2} \right] dz. \quad (18)$$

Since the probability of conserving one of the conditions  $\mu < d/2$  and  $\mu \geq d/2$  is unity, then, in conformance with the theorem of the addition of probabilities

$$P_{n \rightarrow \infty} \left\{ \mu \geq \frac{d}{2} \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp \left[ -\frac{z^2}{2} \right] dz. \quad (19)$$

It was established above that the transmission of information will be without error for  $\mu < d/2$  and can be both with and without error for  $\mu \geq d/2$ . In conformance with this statement and with (19), the probability,  $p_n$ , of the distortion of a code group consisting of  $n$  elements, when automatic correction of the error is used, satisfies the inequality

$$p_n < \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp \left[ -\frac{z^2}{2} \right] dz \quad (20)$$

or

$$p_n < f(p, n, d) \quad (21)$$

where

$$f(p, n, d) = \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp \left[ -\frac{z^2}{2} \right] dz \quad (22)$$

and  $a$  is expressed through (17).

Carrying out calculations with (17) and (22) by making use of existing tables of the right side of (22) which are found in probability texts and putting the results on graphs, we obtain a family of graphs for which the data are given in Table IV.

Using this table when the admissible probability  $p_n$  of the distortion of a code group is given, the code distance  $d$  can be determined which would guarantee this probability for the selected value of the number  $n$  of code group elements and for the known probability  $p$  of the distortion of one element.

Thus, for example, being given  $p_n < 0.01$ ;  $p = 0.1$ , and  $n = 100$ , we obtain  $d \approx 34$ . Here, the auxiliary quantity  $a = 2.33$ .

The (17), (21), and (22), which we found, and the Table IV computed therefrom permit the fundamental parameters of the correcting code to be calculated for sufficiently large  $n$  values. A given noise stability of a coding system will be guaranteed with a certain margin when a code is used with parameters found by such a method. In order to determine the magnitude of this margin, further investigation must be made.

In particular, the  $p_n$  probability of distortion of a code group will be less than 0.01, in the example cited above, for a code distance  $d \geq 34$ . The minimum value of  $d$  guaranteeing a given  $p_n$  probability of code group distortion remains unknown here. It can only be stated that this minimum value is  $d_{\min} \leq 34$ .

ON THE NUMBER OF CODE COMBINATIONS DIFFERING FROM EACH OTHER IN NOT LESS THAN A GIVEN NUMBER OF ELEMENTS (LARGE  $n$  CASE)

One of the most important parameters of a system with a correcting code is the greatest number of code combinations for which any two differ from each other by not

less than a given number of elements. This parameter is expressed by the function  $x_m(n, d)$  with values found in the first section of this work for small  $n$  and  $d$ .

Let us pose the problem of finding the inequality which the  $x_m(n, d)$  function satisfies in the large  $n$  region.

In order to solve this problem, let us select the code spacing

$$d = 2n(p + \beta) \quad (23)$$

where  $\beta$  is any positive constant as small as desired.

Substituting (23) into (17), we find

$$a = \frac{\beta \sqrt{n}}{\sqrt{p(1-p)}}.$$

Here, putting  $n \rightarrow \infty$ , we obtain  $a \rightarrow \infty$ . Returning to (20) and taking into account that its right side approaches zero as  $a \rightarrow \infty$ , we obtain  $p_n \rightarrow 0$ . This means that for a code spacing chosen according to (23), such a sufficiently large number  $n$  of elements in the code group can be found that the  $p_n$  probability of its distortion can be made as small as desired for any value of the  $\beta$  parameter assigned as small as that which one would desire beforehand.

The capacity  $C$  of the system expressed in binary units per unit time is by Shannon's definition:<sup>1</sup>

$$C = \lim_{T \rightarrow \infty} \frac{\log_2 N(T)}{T} \quad (24)$$

where  $N(T)$  is the common number of all different information in time  $T$ .

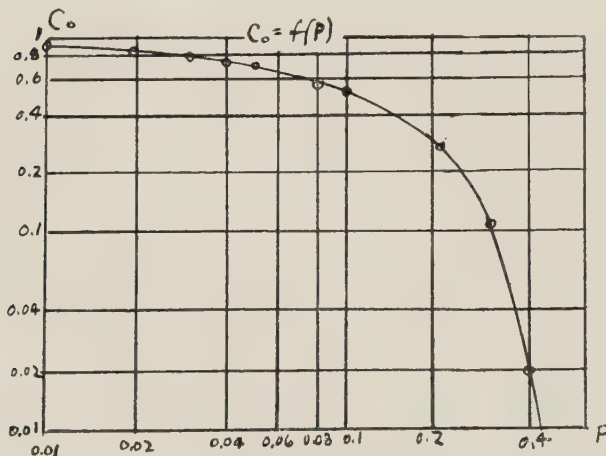


Fig. 4—Dependence of the highest capacity  $C_0$  on the probability  $p$  of damage to one element.

Selecting the length  $\tau$  of the communication element as the time unit (Fig. 4) and taking into account that in the system considered

$$N(T) = x_m(n, d) \quad (25)$$

we obtain

$$C = \lim_{n \rightarrow \infty} \frac{\log_2 x_m(n, d)}{n} \quad (26)$$

According to the above, the channel can be operated at this capacity with the  $p_n$  distortion probabilities as small as desired, if  $n$  is large enough.

Substituting (23) into (26), we obtain

$$C = \lim_{n \rightarrow \infty} \frac{\log_2 x_m(n, 2np + 2n\beta)}{n} \quad (27)$$

On the other hand, it is known that for  $p < 0.5$ , the highest possible capacity  $C_0$  of the system is expressed by Barnard:<sup>7</sup>

$$C_0 = 1 + p \log_2 p + (1 - p) \log_2 (1 - p) \quad (28)$$

which is shown graphically in Fig. 8.<sup>8</sup>

According to the Shannon theorem,<sup>1</sup> such a coding system does not exist which when used would guarantee the transmission of communications at rates exceeding  $C_0$  and with probable errors as close to zero as desired. Consequently

$$C < C_0 \quad (29)$$

or, taking into account (27), we obtain for large enough  $n$

$$\frac{\log_2 x_m(n, 2np + 2n\beta)}{n} < C_0, \quad (30)$$

from which, taking into account that this inequality is correct for  $\beta$  as small as desired, we obtain

$$x_m(n, 2np) < 2^{[1+p \log_2 p + (1-p) \log_2 (1-p)]n}. \quad (31)$$

Here, putting

$$2np = d$$

we find

$$x_m(n, d) < 2^{[1+d/2^n \log_2 d/2^n + (1-d/2^n) \log_2 (1-d/2^n)]n} \quad (32)$$

or, after elementary transformations, we obtain

$$x_m(n, d) < 2^z \quad (33)$$

where

$$z = \phi(\alpha) n$$

$$\phi(\alpha) = \frac{1}{2}[\alpha \log_2 \alpha + (2 - \alpha) \log_2 (2 - \alpha)]$$

$$\alpha = \frac{d}{n}.$$

<sup>7</sup> G. A. Barnard, "Simple Proofs of Simple Cases of the Coding Theorem," paper read at third Symposium on Information Theory 1955.

<sup>8</sup> Translator's note: No Fig. 8 appears in the paper.

Eq. (33) gives the solution of the problem which has been formulated. They permit the determination, for large  $n$  values, of the upper limit for the greatest amount,  $(n, d)$ , of code combinations for which any two are different from each other by not less than a given number of elements.

Shown in Fig. 5 is a graph of  $\phi(\alpha)$  constructed according to (33). Using this graph it is easy to determine the  $z/n$  ratio for a given  $d/n$ .

An analysis of (33) shows that the function  $2^z$ , the upper limit of the desired  $x_m(n, d)$  function, approaches infinity asymptotically for any constant  $d$  and  $n$  increasing without limit. If, conversely,  $n$  remains constant and  $d$  increases from 1 to  $n$ , then the  $2^z$  function decreases monotonically from  $2^n$  to 1. If  $n$  and  $d$  simultaneously increase such that their ratio  $\alpha = d/n$  is constant, then  $\phi(\alpha) = \text{const}$  and the exponent in the  $2^z$  function will increase, approximately, directly proportionally to  $n$ .

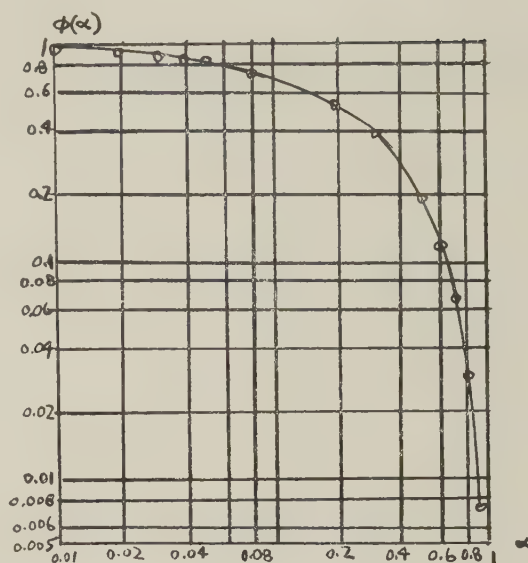


Fig. 5—Graph of  $\phi(\alpha)$ .

## Two Inequalities Implied by Unique Decipherability\*

BROCKWAY McMILLAN†

**Summary**—Consider a list of  $b$  words, each word being a string of letters from a given fixed alphabet of  $a$  letters. If every string of words drawn from this list, when written out in letters without additional space marks to separate the words, is uniquely decipherable, then

$$a^{-l_1} + a^{-l_2} + \cdots + a^{-l_b} \leq 1, \quad (1)$$

where  $l_i$ ,  $1 \leq i \leq b$ , is the length of the  $i$ th word in the list. This result extends a remark of J. L. Doob, who derived the same inequality for lists of a more restricted kind. A consequence of (1) and work of Shannon is that this more restricted kind of list suffices for the search for codes with specified amounts of redundancy.

### DISCUSSION

LET  $A$  be an alphabet of  $a$  letters. A finite nonvacuous sequence or string of letters of  $A$  will be called a *word* over  $A$ . The number of letters in a word will be called its *length*, or length over  $A$ . Consider a finite nonvacuous list  $B$  of words over  $A$ . This list  $B$  will be called *separable* if, whenever a string of words of  $B$  is written out in letters, without space marks between the words, the resulting string of letters is uniquely decipherable into the original string of words. A condition necessary and sufficient for separability is given by Sardinas and Patterson.<sup>1</sup>

The list of common English words, considered as words over the alphabet of 26 Latin letters, is not separable, as the sequences “together” and “to get her” show. If each English word is considered as ending with a space mark, so that the words are over an alphabet of 27 letters, then this list of words is separable.

A strong sufficient condition for separability of  $B$  is that no word of  $B$  appears as the initial string of letters in a longer word of  $B$ . For convenience, call this property of  $B$  *irreducibility*. The list of common English words terminated by spaces is irreducible. The list (1, 10, 100) of words over the binary alphabet is separable but not irreducible.

In an oral discussion,<sup>2</sup> Doob recently observed that the inequality (1) holds when  $B$  is an irreducible list. The main result of this note is a proof of (1) when  $B$  is merely separable. A strong converse result is also given: A construction used by Shannon shows that if  $l_1, l_2, \dots, l_b$  are integers satisfying (1), then there exists an irreducible list  $B$  of  $b$  words whose lengths over  $A$  are respectively  $l_1, l_2, \dots, l_b$ . It follows as a corollary that if  $l_1, l_2, \dots, l_b$  are the lengths over  $A$  of a separable list of words, they are also the lengths over  $A$  of an irreducible list. This rather algebraic conclusion results, as will be seen, from largely analytic arguments.

\* Manuscript received by the PGIT, April 6, 1956.

† Bell Telephone Labs., New York, N. Y.

<sup>1</sup> A. A. Sardinas and G. W. Patterson, “A necessary and sufficient condition for unique decomposition of encoded messages,” 53 IRE CONVENTION RECORD, pt. 8, pp. 104–108.

<sup>2</sup> Comments given upon papers presented before a session on information theory at the summer meeting of the Inst. of Mathematical Statistics, Ann Arbor, Mich., August 30, 1955.

As Doob observed, one interest of (1) is in the following application: Consider a universe of  $b$  events, respectively of probabilities  $p_1, p_2, \dots, p_b$ . Suppose that a code is established in which each event is designated by a distinct word of  $B$ . If the results of a sequence of independent trials are recorded by writing the corresponding words of  $B$  in order without space marks, the expected number of letters written per trial is  $p_1 l_1 + p_2 l_2 + \dots + p_b l_b$ . If (1) holds; *e.g.*, if  $B$  is separable, then one has a direct proof of the inequality

$$\sum_{i=1}^b p_i l_i \geq - \sum_{i=1}^b p_i \log_a p_i. \quad (2)$$

This inequality, of course, also follows from the basic theorems of Shannon.<sup>3</sup>

It follows from the corollary remark above that the set of values assumed by  $\sum p_i l_i$  as  $B$  is varied over the class of separable lists coincides exactly with the set of values assumed when  $B$  is restricted to be irreducible. In particular, given a separable code, there exists an irreducible code for the same universe of events which has the same redundancy.

#### PROOFS

Given a separable list  $B$ , let  $l = \max l_i$  (for  $1 \leq i \leq b$ ), and let  $n_r$  be the number of words of  $B$  which are of length  $r$ ,  $1 \leq r \leq l$ . Then (1) reads

$$\sum_{r=1}^l n_r a^{-r} \leq 1. \quad (1')$$

What will be shown is that the polynomial  $Q(x) - 1$ , where

$$Q(x) = \sum_{r=1}^l n_r x^r$$

has no zeros in the circle  $|x| < a^{-1}$  in the complex plane. In particular then,  $Q(x) - 1$  has no zeros in the interval  $0 \leq x < a^{-1}$ . Since  $Q(x)$  is monotone and continuous for  $x \geq 0$ , and  $Q(0) = 0$ , (1') follows.

Let  $N(k)$  be the number of distinct sequences of words of  $B$  each of which, written as a string of letters, is of total length  $k$  over  $A$ . Since  $B$  is separable, each of these  $N(k)$  sequences of words gives a distinct sequence of  $k$  letters of  $A$ . Hence  $0 \leq N(k) \leq a^k$ . A simple comparison test then shows that for any complex  $x$  such that  $|xa| < 1$  the infinite series  $1 + N(1)x + N(2)x^2 + \dots$  converges. This series, therefore, represents a function  $F(x)$  analytic in  $|xa| < 1$ .

Suppose for a moment that  $k > l$ . Consider the  $N(k)$  strings of  $k$  letters of  $A$  mentioned above: those which are decipherable into sequences of words of  $B$ . They can be

partitioned into  $l$  subclasses  $C_r$ ,  $1 \leq r \leq l$ , thus: Let  $C_r$  consist of all strings whose first word is of length  $r$ . Then if  $r \neq s$ ,  $C_r$  and  $C_s$  cannot overlap; if a string were both, it would be decipherable into two distinct sequences of words. Since there are  $n_r$  distinct words of length  $r$  and  $N(k-r)$  possible distinct subsequent strings of length  $k-r$  which are decipherable into sequences of words of  $B$ ,  $C_r$  contains exactly  $n_r N(k-r)$  distinct strings of letters. Hence if  $k > l$

$$N(k) = n_1 N(k-1) + n_2 N(k-2) + \dots + n_l N(k-l).$$

It is easy to see that if one defines  $N(0) = 1$ , and  $N(-k) = 0$  for  $k > 0$ , then (3) in fact holds for all  $k \geq 0$ .

Multiply (3) by  $x^k$  and sum from  $k = 1$  to  $k = \infty$ . If  $|xa| < 1$ , one gets  $F(x) - 1 = Q(x)F(x)$ , or

$$F(x) = \frac{1}{1 - Q(x)}.$$

Since  $F(x)$  is analytic inside  $|x| < a^{-1}$ ,  $Q(x) - 1$  cannot vanish in that circle. Hence (1').

Conversely, suppose that  $l_1, l_2, \dots, l_b$  are integers satisfying (1). So enumerate them that  $l_1 \leq l_2 \leq \dots \leq l_b$ . It is easy to see that there is then an integer  $k \geq 0$  such that

$$a^{-l_1} + a^{-l_2} + \dots + (k+1)a^{-l_b} = 1.$$

Let  $q_i = a^{-l_i}$ ,  $1 \leq i \leq b-1$ , and  $q_i = a^{-l_b}$  for  $b \leq i \leq b+k$ . Then  $q_1, q_2, \dots, q_{b+k}$  qualify as an exhaustive list of probabilities. The construction of Shannon<sup>3</sup> for an efficient binary code to transmit messages drawn from a universe with the probabilities  $q_i$  can easily be extended to an alphabet  $A$  of  $a$  letters. This extension then describes the construction of an irreducible list of  $b+k$  words, the first  $b$  of which have lengths  $l_1, l_2, \dots, l_b$ . Deleting the last  $k$  words leaves the list irreducible, and with words of the desired length.

Finally to prove (2) from (1), it is necessary only to invoke the inequality

$$\log_a x \leq (x-1) \log_a e, \quad (4)$$

which is well known, and easily proved by elementary calculus. Then if  $p_1, \dots, p_b$  are any nonnegative numbers which sum to 1,

$$\begin{aligned} \sum p_k \log_a \frac{1}{p_k} - \sum p_k l_k &= \sum p_k \log_a \frac{a^{-l_k}}{p_k} \\ &\leq \sum p_k \left( \frac{a^{-l_k}}{p_k} - 1 \right) \log_a e = \end{aligned}$$

Since equality in (4) occurs only when  $x = 1$ , equality (2) can occur only if each  $p_k = a^{-l_k}$ , *i.e.*, if and only if  $l_k = -\log_a p_k$ .

<sup>3</sup> C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379-423, 623-656; July-October, 1948.



# A Note on the Maximum Flow Through a Network\*

P. ELIAS†, A. FEINSTEIN‡, AND C. E. SHANNON§

**Summary**—This note discusses the problem of maximizing the rate of flow from one terminal to another, through a network which consists of a number of branches, each of which has a limited capacity. The main result is a theorem: The maximum possible flow from left to right through a network is equal to the minimum value among all simple cut-sets. This theorem is applied to solve a more general problem, in which a number of input nodes and a number of output nodes are used.

CONSIDER a two-terminal network such as that of Fig. 1. The branches of the network might represent communication channels, or, more generally, any conveying system of limited capacity as, for example, a railroad system, a power feeding system, or a network of pipes, provided in each case it is possible to assign a definite maximum allowed rate of flow over a given branch. The links may be of two types, either one directional (indicated by arrows) or two directional, in which case flow is allowed in either direction at anything up to maximum capacity. At the nodes or junction points of the network, any redistribution of incoming flow into the outgoing flow is allowed, subject only to the restriction of not exceeding in any branch the capacity, and of obeying the Kirchhoff law that the total (algebraic) flow into a node be zero. Note that in the case of information flow, this may require arbitrarily large delays at each node to permit recoding of the output signals from that node. The problem is to evaluate the maximum possible flow through the network as a whole, entering at the left terminal and emerging at the right terminal.

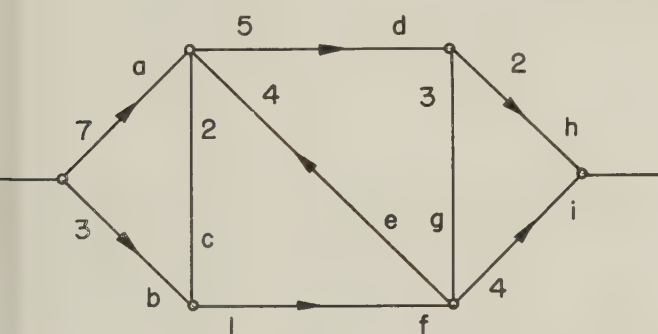


Fig. 1

The answer can be given in terms of cut-sets of the network. A *cut-set* of a two-terminal network is a set of branches such that when deleted from the network, the network falls into two or more unconnected parts with the two terminals in different parts. Thus, every path

from one terminal to the other in the original network passes through at least one branch in the cut-set. In the network above, some examples of cut-sets are  $(d, e, f)$ , and  $(b, c, e, g, h)$ ,  $(d, g, h, i)$ . By a *simple cut-set* we will mean a cut-set such that if any branch is omitted it is no longer a cut-set. Thus  $(d, e, f)$  and  $(b, c, e, g, h)$  are simple cut-sets while  $(d, g, h, i)$  is not. When a simple cut-set is deleted from a connected two-terminal network, the network falls into exactly two parts, a *left part* containing the left terminal and a *right part* containing the right terminal. We assign a *value* to a simple cut-set by taking the sum of capacities of branches in the cut-set, only counting capacities, however, from the left part to the right part for branches that are unidirectional. Note that the direction of an unidirectional branch cannot be deduced from its appearance in the graph of the network. A branch is directed from left to right in a minimal cut-set if, and only if, the arrow on the branch points from a node in the left part of the network to a node in the right part. Thus, in the example, the cut-set  $(d, e, f)$  has the value  $5 + 1 = 6$ , the cut-set  $(b, c, e, g, h)$  has value  $3 + 2 + 3 + 2 = 10$ .

**Theorem:** The maximum possible flow from left to right through a network is equal to the minimum value among all simple cut-sets.

This theorem may appear almost obvious on physical grounds and appears to have been accepted without proof for some time by workers in communication theory. However, while the fact that this flow cannot be exceeded is indeed almost trivial, the fact that it can actually be achieved is by no means obvious. We understand that proofs of the theorem have been given by Ford and Fulkerson<sup>1</sup> and Fulkerson and Dantzig.<sup>2</sup> The following proof is relatively simple, and we believe different in principle.

To prove first that the minimum cut-set flow cannot be exceeded, consider any given flow pattern and a minimum-valued cut-set  $C$ . Take the algebraic sum  $S$  of flows from left to right across this cut-set. This is clearly less than or equal to the value  $V$  of the cut-set, since the latter would result if all paths from left to right in  $C$  were carrying full capacity, and those in the reverse direction were carrying zero. Now add to  $S$  the sum of the algebraic flows into all nodes in the right-hand group for the cut-set  $C$ . This sum is zero because of the Kirchhoff law constraint at each node. Viewed another way, however, we see that it cancels out each flow contributing to  $S$ , and also that each flow on a branch with both ends in the

\* Manuscript received by the PGIT, July 11, 1956.

† Elec. Eng. Dept. and Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

‡ Lincoln Lab., M.I.T., Lexington, Mass.

§ Bell Telephone Labs., Murray Hill, N. J., and M.I.T., Cambridge, Mass.

<sup>1</sup> L. Ford, Jr. and D. R. Fulkerson, *Can. J. Math.*; to be published.

<sup>2</sup> G. B. Dantzig and D. R. Fulkerson, "On the Max-Flow Min-Cut Theorem of Networks," in "Linear Inequalities," *Ann. Math. Studies*, no. 38, Princeton, New Jersey, 1956.

right hand group appears with both plus and minus signs and therefore cancels out. The only term left, therefore, which is not cancelled is the flow out of the right hand terminal, that is to say, the total flow  $F$  through the network. We conclude, then that  $F \leq V$ .

We now prove the more interesting positive assertion of the theorem: That a flow pattern can be found which actually achieves the rate  $V$ . From any given network with minimum cut-set value  $V$  it is possible to construct what we will call a *reduced* network with the properties listed below.

- 1) The graph of the reduced network is the same as that of the original network except possibly that some of the branches of the original network are missing (zero capacity) in the reduced network.
- 2) Every branch in the reduced network has a capacity equal to or less than the corresponding branch of the original network.
- 3) Every branch of the reduced network is in at least one cut-set of value  $V$ , and  $V$  is the minimum value cut-set for the reduced network.

A reduced network may be constructed as follows. If there is any branch which is not in some minimum cut-set, reduce its capacity until either it is in a minimum cut-set or the value reaches zero. Next, take any other branch not in a minimum cut-set and perform the same operation. Continue in this way until no branches remain which are not in minimum cut-sets. The network then clearly satisfies the condition. In general, there will be many different reduced networks obtainable from a given network depending on the order in which the branches are chosen. If a satisfactory flow pattern can be found for a reduced network, it is clear that the same flow pattern will be satisfactory in the original network, since both the Kirchhoff condition and the capacity limitation will be satisfied. Hence, if we prove the theorem for reduced networks, it will be true in general.

The proof will proceed by an induction on the number of branches. First note that if every path through a reduced network contains only two or less elements, the network is of the form shown typically in Fig. 2. In

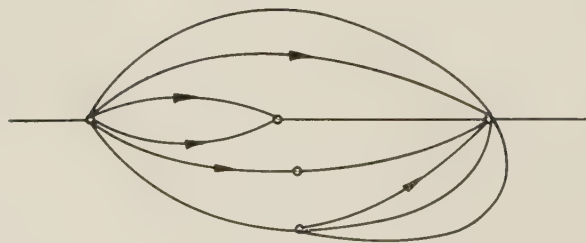


Fig. 2

general, such a network consists of a paralleling of series subnetworks, these series combinations being at most two long with or without arrows from left to right. It is

obvious that for such a reduced network, the theorem is true. It is only necessary to load up each branch to its capacity. Now suppose the theorem true for all reduced networks with less than  $n$  nodes. We will then show that it is true for any reduced network with  $n$  nodes.

Either the given reduced network with  $n$  nodes has a path from left to right of length at least three, or it is the type just described. In the latter case the theorem is true, as mentioned. In the former case, taking the second branch on a path of length three, we have an element running between internal nodes. There exists (since the network is reduced) a minimum cut-set containing this branch. Replace each branch in the cut-set by two branches in series, each with the same capacity as the original branch. Now identify (or join together) all these newly-formed middle nodes as one single node. The network then becomes a series connection of two simpler networks. Each of these has the same minimum cut-set value  $V$  since they each contain a cut-set corresponding to  $C$ , and furthermore neither can contain higher-valued cut-sets since the operation of identifying nodes only eliminates and cannot introduce new cut-sets.

Each of the two networks in series contains a number of branches smaller than  $n$ . This is evident because of the path of length at least three from the left terminal to the right terminal. This path implies the existence of a branch in the left group which does not appear in the right group and conversely. Thus by inductive assumption, a satisfactory flow pattern with total flow  $V$  can be set up in each of these networks. It is clear, then, that when the common connecting node is separated into its original form, the same flow pattern is satisfactory for the original network. This concludes the proof.

It is interesting that in a reduced network each branch is loaded to its full capacity and the direction of flow is determined by any minimum cut-set through a branch. In nonreduced networks there is, in general, some freedom in the amount of flow in branches and even, sometimes, in the direction of flow.

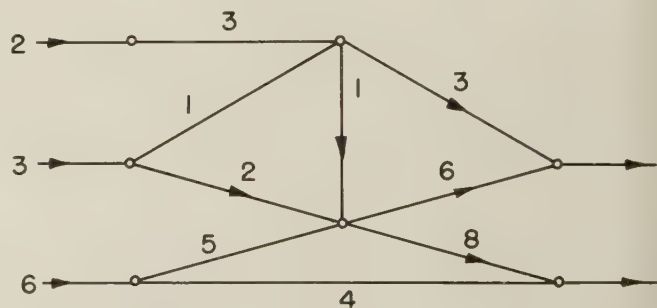


Fig. 3

A more general problem concerning flow through a network can be readily reduced to the above result. Suppose we have a network with a number of input nodes and a number of output nodes as in Fig. 3. The three nodes

on the left are inputs and it is desired to introduce two, three, and six units of flow at these points. The nodes on the right are outputs and it is desired to deliver three and eight units at these points. The problem is to find conditions under which this is possible.

This problem may be reduced to the earlier one by adding a channel for each input to a common left-hand node, the capacity of the channel being equal to the input flow, and also introducing channels from the outputs to a common right-hand node with capacities equal to the output flow. In the particular case this leads to Fig. 4. The

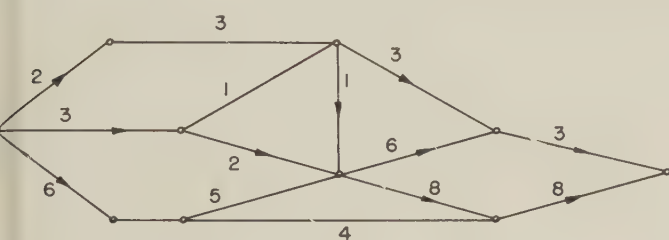


Fig. 4

network obtained in this way from the original problem will be called the *augmented* network.

It is easy to show that necessary and sufficient conditions for solving this multiple input multiple output problem are the following:

- 1) The sum of the input flows must equal the sum of the output flows. Let this sum be  $C$ .
- 2) The minimum cut-set in the augmented network must have a value  $C$ .

To prove these, note that the necessity of 1 is obvious and that of 2 follows by assuming a flow pattern in the original network satisfying the conditions. This can be translated into a flow pattern in the augmented network, and using the theorem, this implies no cut-set with value less than  $C$ . Since there are cut-sets with value  $C$  (those through the added branches), the minimum cut-set value is equal to  $C$ .

The sufficiency of the conditions follows from noting that 2 implies, using the theorem, that a flow pattern can be set up in the augmented network with  $C$  in at the left and out at the right. Now by Kirchhoff's law at the right and left terminals and using condition 1, each added output branch and input branch is carrying a flow equal to that desired. Hence, this flow pattern in the original network solves the problem.

#### ACKNOWLEDGMENT

This work was supported in part by the U. S. Army Signal Corps, the U. S. Air Force Air Research and Development Command, and the U. S. Navy Office of Naval Research.

## Rectification of Two Signals in Random Noise\*

L. LORNE CAMPBELL†

**Summary**—The spectrum of the output of a half-wave rectifier is derived for an input which is the sum of random noise and two sinusoidal signals of different frequencies. The method used is the characteristic function method described by Rice. The components of the output spectrum are given as infinite series of hypergeometric functions. If both the input signals are small compared with the noise, it is shown that the ratio of the output signal power at the difference frequency to the output noise power is proportional to the product of the input signal-to-noise power ratios at the two frequencies. If one of the input signals is very large compared with the noise, it is shown that the other signal and the noise are translated in frequency without alteration of the signal-to-noise ratio. A correction factor is obtained for the case where the large signal is not quite large enough. Finally, the output signal-to-noise ratio of a single-sideband detector is calculated as a function of the input signal-to-noise ratio, when the sideband amplitude is one-half the carrier amplitude.

#### LIST OF PRINCIPAL SYMBOLS

- $P, Q$ —amplitudes of input signals of frequencies  $p/2\pi, q/2\pi$ .  
 $V(t)$ —total input voltage.  
 $V_s(t) = P \cos pt + Q \cos qt$ —total input signal.  
 $V_N(t)$ —input noise voltage.  
 $\alpha, \nu$ —rectifier parameters, defined by  $I = \alpha V^\nu$  for  $V > 0$ ,  $I = 0$  for  $V < 0$  where  $V$  is the input and  $I$  the output.  
 $w(f)$ —power spectrum of input noise.  
 $\psi_0$ —input noise power.  
 $\psi_\tau = \int_0^\infty w(f) \cos 2\pi f\tau df$ —autocorrelation function of input noise.  
 $\Psi(\tau)$ —autocorrelation function of the rectifier output.  
 $\Psi_\infty(\tau)$ —autocorrelation function of the portion of the rectifier output with a discrete spectrum.

\* Manuscript received by the PGIT, May 3, 1956. Work performed under project PCC no. D48-95-12-21.

† Radio Physics Lab., Defence Research Board, Ottawa, Canada.

- $\Psi_c(\tau)$ —autocorrelation function of the portion of the rectifier output with a continuous spectrum.  
 $W(f)$ —power spectrum of rectifier output.  
 $W_c(f)$ —continuous portion of output spectrum.  
 $W_{l.f.}(f)$ —low frequency output noise spectrum.  
 $g(u, v, \tau)$ —characteristic function of the probability density of  $V(t), V(t + \tau)$ .  
 $g_s(u, v, \tau)$ —characteristic function of  $V_s(t), V_s(t + \tau)$ .  
 $x = P^2/2\psi_0, y = Q^2/2\psi_0$ —input signal-to-noise power ratios for the two input signals.  
 $k, m, n, s$ —integers.  
 $C$ —contour along real axis with downward indentation at the origin.  
 $G_k(f) = \int_0^\infty (\psi_\tau)^k \cos 2\pi f\tau d\tau$ .  
 $r_{mnk}$ —amplitudes of various spectral components in output.  
 $\epsilon_0 = 1, \epsilon_n = 2$  for  $n > 0$ .  
 $J_m(x)$ —Bessel function of order  $m$ .  
 $\Gamma(x)$ —Gamma function.  
 ${}_1F_1(a; c; x)$ —confluent hypergeometric function.  
 ${}_2F_1(a, b; c; x)$ —hypergeometric series.  
 $f_a = |p - q|/2\pi$ —the difference frequency of the two signals.  
 $I_s$ —output signal at frequency  $f_a$ .  
 $\beta$ —input bandwidth.  
 $\Delta f$ —output bandwidth.  
 $\rho$ —output signal-to-noise power ratio in the band  $\Delta f$ .  
 $w_0$ —input noise power per unit bandwidth when input noise spectrum is rectangular.  
 $\lambda$ —modulation index of single-sideband signal, Section VI only.  
 $\omega/2\pi$ —carrier frequency of single-sideband signal.  
 $\Delta\omega/2\pi$ —modulation frequency of single-sideband signal.

## I. INTRODUCTION

THE SPECTRUM of the output of a  $\nu$ -th law half-wave rectifier when the input is a sinusoidal signal combined with Gaussian noise has been determined by Rice<sup>1</sup> and Middleton.<sup>2,3</sup> Middleton also gives the output spectrum when the input is an amplitude-modulated signal combined with noise. Rice gives the output from a full-wave square law rectifier when the input is noise plus a sinusoidal signal, the sum of two sinusoidal signals, or an amplitude-modulated signal.

In this paper we derive the spectrum of the output from a  $\nu$ -th law half-wave rectifier when the input is the sum of two sinusoidal signals of different frequencies and Gaussian noise. Thus, let the input voltage be

$$V(t) = P \cos pt + Q \cos qt + V_N(t), \quad (1)$$

where  $V_N$  is a noise voltage. We derive the power spectrum of  $I(t)$  where

$$I(t) = \alpha[V(t)]^\nu \quad \text{for } V(t) \geq 0 \\ = 0 \quad \text{for } V(t) < 0. \quad (2)$$

General expressions will be obtained for arbitrary  $\nu$  and some special cases will be treated in more detail for the linear half-wave rectifier ( $\nu = 1$ ).

An input voltage of the form (1) may be found in such devices as single-sideband detectors and mixers, and in some forms of phase detector. The noiseless case has been treated theoretically by Bennett<sup>4</sup> for half-wave linear and square law rectifiers. The resulting modulation products for these cases have been tabulated by Sternberg, Shipman, Thurston, and Kaufman.<sup>5,6</sup>

The power spectrum of the output of the rectifier derived in Section II and in Section III these results are put into a form which is more useful for calculation. In Sections IV, V, and VI the special cases of small signals, large signals, and single-sideband detection are considered.

## II. DERIVATION OF THE OUTPUT SPECTRUM

We use the characteristic function method described by Rice.<sup>7</sup> He has already derived the dc output and the total low-frequency output by a slightly different method for the case of a linear rectifier.

The autocorrelation function,  $\Psi(\tau)$ , of the output of the rectifier is given<sup>7</sup> by

$$\Psi(\tau) = \frac{\alpha^2}{4\pi^2} \int_C \frac{\Gamma(\nu+1) du}{(iu)^{\nu+1}} \int_C \frac{\Gamma(\nu+1) dv}{(iv)^{\nu+1}} g(u, v, \tau) dv, \quad (3)$$

where

$$g(u, v, \tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \exp[iuV(t) + ivV(t + \tau)] dt, \quad (4)$$

and the contour  $C$  is the real axis from  $-\infty$  to  $\infty$  with downward indentation at the origin. The output power spectrum,  $W(f)$ , of  $\Psi(\tau)$  is given by the transform

$$W(f) = 4 \int_0^\infty \Psi(\tau) \cos 2\pi f\tau d\tau. \quad (5)$$

If  $V_N$  represents Gaussian noise, the characteristic function,  $g(u, v, \tau)$ , is given by

$$g(u, v, \tau) = g_s(u, v, \tau) \exp \left[ -\frac{\psi_0}{2} (u^2 + v^2) - \psi_c uv \right], \quad (6)$$

<sup>4</sup> W. R. Bennett, "New results in the calculation of modulation products," *Bell Sys. Tech. J.*, vol. 12, pp. 228-243; April, 1933.

<sup>5</sup> R. L. Sternberg, J. S. Shipman, and W. B. Thurston, "Tables of Bennett functions for the two-frequency modulation product problem for the half-wave linear rectifier," *Quart. J. Mech. and Appl. Math.*, vol. 7, pp. 505-511; December, 1954.

<sup>6</sup> R. L. Sternberg, J. S. Shipman, and H. Kaufman, "Tables of Bennett functions for the two-frequency modulation product problem for the half-wave square-law rectifier," *Quart. J. Mech. and Appl. Math.*, vol. 8, pp. 457-467; December, 1955.

<sup>7</sup> S. O. Rice, *op. cit.*, sections 4.8 to 4.10.

<sup>8</sup> *Ibid*, section 4.2.

<sup>1</sup> S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46-156; January, 1945.

<sup>2</sup> D. Middleton, "Rectification of a sinusoidally modulated carrier in the presence of noise," *Proc. IRE*, vol. 36, pp. 1467-1477; December, 1948.

<sup>3</sup> D. Middleton, "Some general results in the theory of noise through non-linear devices," *Quart. Appl. Math.*, vol. 5, pp. 445-498; January, 1948.

where  $\psi_\tau$  is the autocorrelation function of the noise,  $\psi_0$  is the mean square value of the noise

$$g_s(u, v, \tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \exp [iuV_s(t) + ivV_s(t + \tau)] dt \quad (7)$$

and in the case of the input (1)

$$V_s(t) = P \cos pt + Q \cos qt. \quad (8)$$

Now, it is easily shown from the generating function for Bessel functions that

$$\exp(iuP \cos pt) = \sum_{n=0}^{\infty} i^n \epsilon_n J_n(Pu) \cos npt, \quad (9)$$

where  $\epsilon_0 = 1$  and  $\epsilon_n = 2$  for  $n > 0$ . Thus, in order to evaluate the integral of (7), it will be seen that we must consider integrals of the form

$$I_{n_1 n_2 n_3 n_4} = \int_0^T \cos n_1 pt \cos n_2 qt \cdot \cos n_3 p(t + \tau) \cos n_4 q(t + \tau) dt. \quad (10)$$

It will be assumed that  $p$  and  $q$  are incommensurable. That is, it is assumed that  $mp \pm nq \neq 0$  for any integers  $m$  and  $n$ . Then when we form

$$\lim_{T \rightarrow \infty} \frac{1}{T} I_{n_1 n_2 n_3 n_4},$$

the only terms which contribute are those for which  $n_1 = n_3$  and  $n_2 = n_4$ , since all other integrals contain only terms like  $\cos[(n_3 \pm n_1)pt + n_3 p\tau]$  and  $\cos[(n_4 \pm n_2)qt + n_4 q\tau]$ .

Now

$$I_{n_1 n_2 n_1 n_2} = \frac{1}{4} \int_0^T [\cos(2n_1 pt + n_1 p\tau) + \cos n_1 p\tau] \cdot [\cos(2n_2 qt + n_2 q\tau) + \cos n_2 q\tau] dt, \quad (11)$$

and hence

$$\lim_{T \rightarrow \infty} \frac{1}{T} I_{n_1 n_2 n_1 n_2} = \frac{1}{\epsilon_{n_1} \epsilon_{n_2}} \cos n_1 p\tau \cos n_2 q\tau. \quad (12)$$

Therefore we have

$$g_s(u, v, \tau) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \epsilon_m \epsilon_n J_m(Pu) J_m(Pv) J_n(Qv) J_n(Qv) \cdot \cos mp\tau \cos nq\tau. \quad (13)$$

In order to evaluate (3) we expand  $e^{-\psi_\tau uv}$  in a power series. We have

$$e^{-\psi_\tau uv} g_s(u, v, \tau) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{k+m+n} \epsilon_m \epsilon_n J_m(Pu) J_n(Qv) \cdot J_m(Pv) J_n(Qv) \frac{(\psi_\tau uv)^k}{k!} \cos mp\tau \cos nq\tau. \quad (14)$$

Therefore, from (3), (6), and (14),

$$\Psi(\tau) = \alpha^2 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \epsilon_m \epsilon_n \cos mp\tau \cos nq\tau \frac{(\psi_\tau)^k}{k!} r_{mnk}^2, \quad (15)$$

where

$$r_{mnk} = i^{k+m+n-\nu-1} \frac{\Gamma(\nu+1)}{2\pi} \cdot \int_C e^{-\psi_0 u^2/2} J_m(Pu) J_n(Qv) u^{k-\nu-1} du. \quad (16)$$

The output spectrum is given by (5). It is convenient to separate the discrete spectrum [dc component and components of frequency  $(mp \pm nq)/2\pi$ ] from the continuous spectrum. The discrete spectrum is obtained from the terms in (15) which do not vanish as  $\tau \rightarrow \infty$ . These are the terms for which  $k = 0$ , since  $\psi_\tau \rightarrow 0$  as  $\tau \rightarrow \infty$ . Thus we shall write

$$\Psi(\tau) = \Psi_\infty(\tau) + \Psi_c(\tau), \quad (17)$$

where

$$\Psi_\infty(\tau) = \alpha^2 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_m \epsilon_n r_{mn0}^2 \cos mp\tau \cos nq\tau, \quad (18)$$

and  $\Psi_c(\tau)$  is the portion of  $\Psi(\tau)$  which contributes to the continuous spectrum.

Now,

$$\Psi_\infty(\tau) = \frac{\alpha^2}{2} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_m \epsilon_n r_{mn0}^2 [\cos(mp + nq)\tau + \cos(mp - nq)\tau]. \quad (19)$$

Thus the amplitude of the dc component of the output,  $I$ , is

$$I_{D.C.} = \alpha |r_{000}| \quad (20)$$

and the amplitude of the components of frequencies  $(mp + nq)/2\pi$  and  $|mp - nq|/2\pi$  are

$$I_{mp \pm nq} = 2\alpha |r_{mn0}|. \quad (21)$$

The continuous spectrum,  $W_c(f)$ , is obtained from

$$W_c(f) = 4 \int_0^\infty \Psi_c(\tau) \cos 2\pi f\tau d\tau. \quad (5')$$

Therefore,

$$W_c(f) = \alpha^2 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \epsilon_m \epsilon_n \frac{r_{mnk}^2}{k!} \left[ G_k\left(\frac{mp + nq}{2\pi} + f\right) + G_k\left(\frac{mp + nq}{2\pi} - f\right) + G_k\left(\frac{mp - nq}{2\pi} + f\right) + G_k\left(\frac{mp - nq}{2\pi} - f\right) \right], \quad (22)$$

where

$$G_k(f) = \int_0^\infty (\psi_\tau)^k \cos 2\pi f\tau d\tau. \quad (23)$$

The functions  $G_k(f)$  are discussed and tabulated by Rice<sup>9</sup> for three types of predetector filter.

### III. EVALUATION OF THE CONSTANTS $r_{mnk}$

It is necessary to obtain a more useful expression for  $r_{mnk}$  from (16). Two infinite series expressions for  $r_{mnk}$  will be derived here.

The first expression is obtained by expanding  $J_n(Qu)$  in a Maclaurin series. We obtain

$$r_{mnk} = i^{k+m+n-\nu-1} \frac{\Gamma(\nu+1)}{2\pi} \sum_{s=0}^{\infty} \frac{(-1)^s (Q/2)^{n+2s}}{s!(s+n)!} \cdot \int_C e^{-\psi_0 u^2/2} J_m(Pu) u^{k+n+2s-\nu-1} du. \quad (24)$$

The integrals in (24) may be evaluated by means of a formula given by Rice.<sup>7</sup> The result is

$$r_{mnk} = \frac{\Gamma(\nu+1)}{2m!} \left(\frac{2}{\psi_0}\right)^{(k-\nu)/2} x^{m/2} y^{n/2} \cdot \sum_{s=0}^{\infty} \frac{y^s {}_1F_1(1/2[k+m+n+2s-\nu]; m+1; -x)}{s!(s+n)! \Gamma(1/2[2-k-m-n-2s+\nu])} \quad (25)$$

where

$$x = \frac{P^2}{2\psi_0}, \quad y = \frac{Q^2}{2\psi_0}. \quad (26)$$

The quantities  $x$  and  $y$  are signal-to-noise power ratios at the input for the two signals  $P \cos pt$  and  $Q \cos qt$ . The function  ${}_1F_1(a; c; z)$  is a confluent hypergeometric function, defined by

$${}_1F_1(a; c; z) = 1 + \frac{az}{c} + \frac{a(a+1)}{c(c+1)} \frac{z^2}{2!} + \dots \quad (27)$$

It may be noted that, if  $k+m+n-\nu$  is an even integer, the series in (25) terminates because of the Gamma function in the denominator.

An alternative series for  $r_{mnk}$  may be obtained by expanding the product  $J_m(Pu) J_n(Qu)$  as follows.<sup>10</sup>

$$J_m(Pu) J_n(Qu) = \frac{(Pu/2)^m (Qu/2)^n}{n!} \cdot \sum_{s=0}^{\infty} \frac{(-1)^s (Pu/2)^{2s} {}_2F_1(-s, -m-s; n+1; Q^2/P^2)}{s!(m+s)!}, \quad (28)$$

where  ${}_2F_1(a, b; c; z)$  is the hypergeometric series, given by

$${}_2F_1(a, b; c; z) = 1 + \frac{abz}{c} + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{z^2}{2!} + \dots \quad (29)$$

The series (29) terminates if  $a$  or  $b$  is a negative integer. If we insert the series (28) in (16) and use the result<sup>7</sup> that

$$\int_C e^{-au^2} u^{2\lambda-1} du = \frac{i\pi e^{-\lambda i\pi}}{a^\lambda \Gamma(1-\lambda)}, \quad (30)$$

we obtain

$$r_{mnk} = \frac{\Gamma(\nu+1)}{2n!} x^{m/2} y^{n/2} \left(\frac{2}{\psi_0}\right)^{(k-\nu)/2} \cdot \sum_{s=0}^{\infty} \frac{{}_2F_1(-s, -m-s; n+1; y/x) x^s}{s!(m+s)! \Gamma(1/2[2-k-m-n-2s+\nu])}. \quad (31)$$

Each  ${}_2F_1$  function in (31) is a finite polynomial in  $y/x$ .

### IV. SMALL INPUT SIGNAL-TO-NOISE RATIO

In this section and the succeeding sections we shall assume that the desired output is the signal with frequency  $f_a = |p - q|/2\pi$  and that post-detection filters remove all the other signals and all the noise except that in a band around the desired signal. Then the output signal,  $I_s$ , is given by

$$I_s = 2\alpha r_{110}. \quad (32)$$

We shall also assume that the noise input,  $V_N$ , has a rectangular spectrum given by

$$w(f) = w_0 = \frac{\psi_0}{\beta} \quad f_0 - \frac{\beta}{2} < |f| < f_0 + \frac{\beta}{2} \quad (33)$$

$$w(f) = 0 \quad \text{elsewhere,}$$

where the frequencies  $p/2\pi$  and  $q/2\pi$  are between  $f_0 - \beta/2$  and  $f_0 + \beta/2$ . Then, as shown by Rice,<sup>9</sup>

$$G_1(f) = \frac{\psi_0}{4\beta} \quad f_0 - \frac{\beta}{2} < |f| < f_0 + \frac{\beta}{2} \quad (34)$$

$$G_1(f) = 0 \quad \text{elsewhere,}$$

and the low-frequency portion of  $G_2(f)$  is given by

$$G_2(f) = \frac{\psi_0^2}{4\beta} (1 - f/\beta) \quad 0 \leq f \leq \beta$$

$$G_2(f) = 0 \quad |f| > \beta. \quad (35)$$

We shall consider a relatively narrow-band system; that is, we assume  $\beta \ll p$ ,  $|p - q| \ll p$ .

In this section we consider the case where the input signal-to-noise ratios  $x$  and  $y$  are both small compared with unity. The output signal-to-noise ratio in a band of width  $\Delta f$  about the signal will be obtained here.

From (25) it is seen that, for  $x \ll 1$  and  $y \ll 1$ , the signal amplitude is given approximately by

$$I_s = 2\alpha r_{110} = \frac{\alpha \Gamma(\nu+1)}{\Gamma(\nu/2)} \left(\frac{\psi_0}{2}\right)^{\nu/2} (xy)^{1/2}. \quad (36)$$

The principal contribution to the low-frequency noise comes from the beats of noise with itself, *i.e.*, from the term in (22) involving  $r_{002}$ . The noise power in the band  $\Delta f$  is given approximately by

$$N = \alpha^2 r_{002}^2 \int_{f_a - \Delta f/2}^{f_a + \Delta f/2} [G_2(f) + G_2(-f)] df, \quad (37)$$

or

$$N = \frac{\alpha^2}{2} \left[ \frac{\Gamma(\nu+1)}{\Gamma(\nu/2)} \right]^2 \left(\frac{\psi_0}{2}\right)^\nu \frac{\Delta f}{\beta} \left(1 - \frac{f_a}{\beta}\right). \quad (38)$$

<sup>9</sup> *Ibid.*, appendix.

<sup>10</sup> G. N. Watson, "Theory of Bessel Functions," The Macmillan Co., New York, N. Y., p. 148; 1948.

The output signal-to-noise power ratio,  $\rho$ , is given by

$$\rho = \frac{I_s^2}{2N} = \frac{\beta xy}{\Delta f(1 - f_a/\beta)}. \quad (39)$$

or  $f_a < \beta$ . This result is analogous to the result obtained by Middleton<sup>2</sup> for the rectification of amplitude modulated waves, *viz.*, that the output signal-to-noise ratio is proportional to the square of the input signal-to-noise ratio for small values of the input signal-to-noise ratio. This result is independent of the exponent  $\nu$  in the detector law.

## V. ONE LARGE INPUT SIGNAL—LINEAR RECTIFIER

In this section we consider the output spectrum of a linear ( $\nu = 1$ ) rectifier when  $x \gg 1$  and  $x \gg y$ . This is the condition which normally holds if the signal  $P \cos pt$  is supplied by a local oscillator. For this purpose the asymptotic relation<sup>9</sup>

$$F_1(\alpha; \beta; -p) \simeq p^{-\alpha} \frac{\Gamma(\beta)}{\Gamma(\beta - \alpha)} \left[ 1 + \frac{\alpha(\alpha - \beta + 1)}{p} + \frac{\alpha(\alpha + 1)(\alpha - \beta + 1)(\alpha - \beta + 2)}{p^2 2!} + \dots \right] \quad (40)$$

( $p > 0$ )

is useful.

From (25) and (40) it is seen that as  $x \rightarrow \infty$ , the terms of highest order in the spectrum are those proportional to  $r_{m00}^2$  ( $m = 0, 1, 2, \dots$ ). These terms are just the dc and the harmonics of  $P \cos pt$ . As a check on the results, it is easily shown that the amplitudes of the dc and harmonics agree with those calculated from the Fourier series of the output with  $Q$  and  $V_N$  equal to zero.

Finally, we shall consider the low-frequency spectrum when  $x$  is large. Then

$$r_{110} \simeq \frac{Q}{2\pi} \left[ 1 - \frac{3}{8x} + O(x^{-2}) \right], \quad (41)$$

so that

$$I_s = 2\alpha r_{110} \simeq \alpha \frac{Q}{\pi} \left[ 1 - \frac{3}{8x} + O(x^{-2}) \right]. \quad (42)$$

The principal contribution to the low-frequency noise spectrum (39) is from  $r_{101}$ . For  $x \gg 1$ ,

$$r_{101} \simeq \frac{1}{\pi} \left[ 1 - \frac{1}{2x} + O(x^{-2}) \right]. \quad (43)$$

All other contributions are of order  $x^{-1/2}$  or smaller. Now, if  $w(f)$  is the input spectrum, it is seen easily from (5) and (23) that

$$G_1(f) = \frac{w(f)}{4}. \quad (44)$$

Thus, the low-frequency noise spectrum is, approximately,

$$W_{L.F.}(f) = \alpha^2 r_{101}^2 \left[ w\left(f + \frac{p}{2\pi}\right) + w\left(f - \frac{p}{2\pi}\right) \right]. \quad (45)$$

It will be seen that the noise spectrum has been split into two parts, which have been shifted up and down, respectively, in frequency by an amount  $p/2\pi$ . The noise power spectrum has, from (43), been multiplied also by a factor  $\alpha^2/\pi^2$  for large  $x$ . At the same time the signal amplitude,  $I_s$ , has been multiplied by a factor  $\alpha/\pi$  so that the signal power is multiplied by  $\alpha^2/\pi^2$ .

Thus, when  $x$  is large enough, the effect on  $Q \cos qt + V_N$  is just to shift the frequency spectrum of this voltage without altering the shape of the spectrum or the relation between signal and noise.

When  $P$  is not quite large enough to shift  $Q \cos qt + V_N$  in frequency without distortion, a more nearly exact expression can be obtained by including more terms in the asymptotic expansions. For example, consider the case where the input spectrum is rectangular with a width which is large compared with  $|p - q|/2\pi$ . For convenience we will assume that the input noise has a spectrum given by (33), with the center frequency  $f_0$  equal to  $p/2\pi$  and with  $\beta \gg |p - q|/2\pi$ . In this case, the low-frequency noise in the output in a band of width  $\Delta f$  about the signal frequency is given approximately by

$$N = \alpha^2 \Delta f \left[ 8r_{101}^2 G_1\left(\frac{p}{2\pi}\right) + 8r_{011}^2 G_1\left(\frac{q}{2\pi}\right) + 2r_{002}^2 G_2(0) \right], \quad (46)$$

or

$$N = \frac{\alpha^2 \psi_0 \Delta f}{2\beta} [4r_{101}^2 + 4r_{011}^2 + \psi_0 r_{002}^2]. \quad (47)$$

If the asymptotic expansion (40) is used to calculate the coefficients  $r_{101}$ ,  $r_{011}$ , and  $r_{002}$  an approximate expression can be obtained for the output noise. Then the output signal-to-noise ratio,  $\rho$ , is given by

$$\rho = \frac{I_s^2}{2N} \simeq \frac{\beta}{2\Delta f} y \left( 1 - \frac{1}{8x} \right). \quad (48)$$

According to (48) the output signal-to-noise ratio,  $\rho$ , is just equal to the input signal-to-noise ratio,  $y$ , multiplied by the bandwidth factor  $\beta/2\Delta f$ , provided that  $P$ , and hence  $x$ , is large enough. If  $P$  is not quite large enough to transfer the signal,  $Q \cos qt$ , in frequency without altering the signal-to-noise ratio, the first-order correction factor is  $1 - 1/8x$ .

## VI. SINGLE-SIDEBAND RECEPTION—LINEAR RECTIFIER

In this section we consider the detection of the signal

$$V = E_0 [\cos \omega t + \lambda \cos (\omega + \Delta\omega)t] + V_N \quad (49)$$

in a linear rectifier. Bridges<sup>11</sup> has treated a similar problem by a different method. It will be assumed that  $\Delta\omega \ll \omega$ . Such a signal could arise in single-sideband transmission with unsuppressed or partially suppressed carrier.  $E_0 \cos \omega t$  will be referred to here as the carrier and  $x$  will be the carrier to noise ratio, given by

<sup>11</sup> J. E. Bridges, "Detection of television signals in thermal noise," *Proc. IRE*, vol. 42, pp. 1396-1405; September, 1954.

$$x = \frac{E_0^2}{2\psi_0}. \quad (50)$$

Then, in the notation used earlier,

$$y = \frac{E_0^2 \lambda^2}{2\psi_0} = \lambda^2 x. \quad (51)$$

Only the terms in the output low frequency noise spectrum which involve the coefficients  $r_{002}$ ,  $r_{101}$ , and  $r_{011}$  will be considered here. These terms represent, respectively, the contributions due to beats of noise with noise, noise with carrier, and noise with sideband. The approximation of the series in (22) by three terms is equivalent to the approximation made by Middleton<sup>2</sup> in his calculation for double-sideband amplitude modulation. It will be assumed that the width of the input noise spectrum is large compared with the modulation frequency  $\Delta\omega/2\pi$  and that the signal is near the center of the input spectrum. Then the output noise power in a band  $\Delta f$  about the signal is given approximately by (47) with  $y$  given by (51). Hence the output signal-to-noise ratio is given by

$$\rho = \frac{4\beta r_{110}^2}{\psi_0 \Delta f [\psi_0^2 r_{002}^2 + 4r_{101}^2 + 4r_{011}^2]}. \quad (52)$$

The coefficients  $r_{mnk}$  were determined as functions of  $x$  for  $\lambda = 1/2$  with the aid of (25). The confluent hypergeometric functions were obtained from tables prepared by Middleton and Johnson.<sup>12</sup> The output signal-to-noise ratio,  $\rho$ , is plotted in Fig. 1 as a function of carrier-to-noise ratio,  $x$ , for  $\lambda = 1/2$ . The units of  $\rho$  are  $\beta/2\Delta f$  so that the signal-to-noise ratio shown must be multiplied by the bandwidth factor  $\beta/2\Delta f$ .

The form of single-sideband reception in which the carrier is supplied by a local oscillator follows easily from the results of Section V.

## VII. DISCUSSION OF RESULTS

The formulas of Sections II and III may provide useful methods of computing the output of a half-wave rectifier when the input consists of the sum of two sinusoidal signals and random noise. The usefulness of these formulas depends, in part, on the availability of tables of the confluent hypergeometric functions. Almost all the functions which occur in (25) are of the form  ${}_1F_1(M/2; N; -x)$ , where  $M$  is an odd integer and  $N$  is an integer.

<sup>12</sup> D. Middleton and V. Johnson, "A Tabulation of Selected Confluent Hypergeometric Functions," Cruft Lab. Tech. Rep. no. 140; January, 1952.

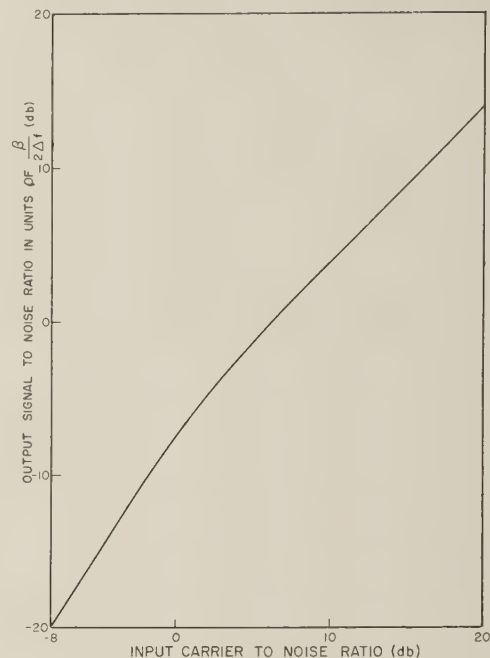


Fig. 1—Output signal-to-noise ratio as a function of input carrier-to-noise ratio for single-sideband detection with  $\lambda = 1/2$ .

The terms for which  $M$  is an even integer drop out, in almost every case, because of a Gamma function with a nonpositive integer argument in the denominator. As mentioned before, Middleton and Johnson<sup>12</sup> have tabulated many confluent hypergeometric functions of the required form. It may also be mentioned that  ${}_1F_1(M/2; N; -x)$  can be expressed in terms of the modified Bessel functions  $I_0(x/2)$  and  $I_1(x/2)$ . Tables of these Bessel functions are generally available.

The result of Section IV agrees with the result obtained for other signals. The result is that a half-wave detector seriously degrades the signal-to-noise ratio if the input signal-to-noise ratio is too low. This is the familiar phenomenon of modulation suppression.

Section V yields the well-known result that a signal and noise can be translated in frequency without modifying the relationship between the signal and noise provided that the oscillator signal is large enough. If the signal provided by the local oscillator is not quite large enough to do this, (48) may be used to obtain a first-order approximation to the signal-to-noise ratio at the new frequency. Better approximations could be obtained by the use of more terms in the asymptotic expansions.

Section VI provides one example of the type of calculation which must be performed when it is not possible to use the approximations of Sections IV or V.



# Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication, I\*

ROBERT PRICE†

**Summary**—Solutions are obtained in open form for the optimum, probability-computing detector of either Gaussian signals, or known signals transmitted via scatter-paths, where the signals have been further perturbed by additive white Gaussian noise. The optimum receiver operates on the received waveforms with filter-functions and biasing constants determined by pairs of inhomogeneous and homogeneous integral equations, respectively.

General solution in closed form has not been obtained, but it is possible to draw a few broad conclusions, among them that the filter-functions can be physically realizable. Approximate solution for the optimum scatter-path receiver) at small signal-to-noise ratios yields a block diagram having interesting implications. For a single-scatter-path, the optimum receiver may be interpreted as the combination of a correlator with an optimum estimator of the Wiener type. Certain special cases in which complete solution is possible have been investigated in detail, and appropriate curves are presented.

The role and performance of the probability-computing detector in an optimum decision-making receiver, for the types of channels considered, is deferred to a companion paper.

## INTRODUCTION

THE PROBLEM of detecting the presence or absence of a Gaussian signal transmitted through a channel containing additive Gaussian noise has received attention from several authors.<sup>1-6</sup> Relatively few explicit results, however, appear to have been obtained on the best detection procedure for such random transmitted signals, in comparison with similar studies<sup>5,7,8</sup> that have dealt with the detection of transmitted signals whose waveforms are completely known to the receiver.

Heaps' analysis<sup>4</sup> for Gaussian signals in Gaussian

noise, an interesting contribution to noise theory, does not appear applicable to the optimum detection problem. Kaplan's results<sup>3</sup> apply to a particular detection scheme which is, in general, not the best one attainable. In the work of Peterson, Birdsall, and Fox,<sup>5</sup> a best detection scheme is derived for the very special case in which signals and noise have rectangular power spectra of equal width.

It appears to be only in the work of R. C. Davis,<sup>2</sup> based on Grenander's studies, that a completely rigorous and general approach to the problem of optimum detection<sup>9</sup> of Gaussian signals in Gaussian noise has been formulated. Here, however, practically no results have been obtained in closed form, even in the somewhat unrealistic case of the complete statistical specification of noise and signal.

The main purpose of this paper is to present functional forms for optimum detectors which, in general, are more explicit than those of Davis, and which, in some special cases, can be specified in terms of operations carried out on the received signal  $w(t)$  by filters of completely determined characteristics. Such extension of Davis' work, however, has been made possible only by making certain simplifying assumptions at the start, such as requiring complete specification of all statistical parameters and

<sup>9</sup> In order to define the "optimum" detector, we assume a communication system model in which the transmitter sends through a noisy channel a waveform  $z^{(k)}(t)$  drawn at random from one of a finite number  $M$  of distinct ensembles, each ensemble having complete statistical specification. The generating ensemble is itself selected according to a symbol  $s_k$ ,  $k = 1, 2, \dots, M$ , originating in an information source at the transmitter, with *a priori* probability  $P_k$ . The detector observes the received wave  $w(t)$  over a time interval  $T$  and takes into account *a priori* knowledge of the  $P_k$  and the statistical parameters of the  $M$  ensembles and the channel noise. Using this data, the detector yields information concerning which of the  $s_k$  was encoded at the transmitter. We distinguish between two kinds of optimum detectors:

1) If it is not required that the detector make a decision, or "guess," no information about the  $s_k$  is lost, according to the philosophy of P. M. Woodward and I. L. Davies, "Information theory and inverse probability in telecommunications," *Proc. IEE*, vol. 99, part III, pp. 37-44; March, 1952, if the detector output consists of the *a posteriori* conditional probabilities  $P_T[s_k/w(t)]$ . (The subscript  $T$  indicates a limited observation interval; it will be dropped when the observation interval is allowed to extend over an infinite interval.) Accordingly, a detector that computes these probabilities will be defined as optimum in this case.

2) If decision is required, an optimum detector is one in which the ill effects resulting from a "wrong guess" would be a minimum. For example, in the latter half of this paper we are concerned with the decision-detector that yields minimum probability of error, Siebert's Ideal Observer. This is only one detector, however, in the broad class of optimum decision-detectors that arises from Decision Theory, a topic which has received extensive and rewarding investigation by Middleton and Van Meter, *loc. cit.*, and Peterson, Birdsall, and Fox, *loc. cit.* It is beyond the scope of this paper to review these many important contributions but we may remark that one most significant result of these studies is that, when available, the computed  $P_T[s_k/w(t)]$  govern the decision in all such optimum decision-detectors.

\* Manuscript received by the PGIT, August 31, 1956. The research in this paper was supported jointly by the Army, Navy, and Air Force under contract with Mass. Inst. Tech., Cambridge, Mass.

† Lincoln Lab., M.I.T., Cambridge, Mass.

<sup>1</sup> D. Middleton, "On the detection of stochastic signals in additive normal noise, I," to be published in this journal at a future date.

<sup>2</sup> R. C. Davis, "The detectability of random signals in the presence of noise," *IRE TRANS.*, PGIT-3, pp. 52-62; March, 1954.

<sup>3</sup> E. L. Kaplan, "Signal detection studies, with applications," *Bell Sys. Tech. J.*, vol. 34, pp. 403-437; March, 1955.

<sup>4</sup> H. S. Heaps, "The effect of a random noise background upon the detection of a random signal," *Can. J. Phys.*, vol. 33, pp. 1-10; January, 1955.

<sup>5</sup> W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The theory of signal detectability," *IRE TRANS.*, PGIT-4, pp. 171-212; September, 1954. See sec. 4.6.

<sup>6</sup> U. Grenander, "Stochastic processes and statistical inference," *Arkiv Mat.*, Band 1, Häfte 3, 1950.

<sup>7</sup> D. Middleton and D. Van Meter, "Detection and extraction of signals in noise from the point of view of statistical decision theory," *J. Soc. Indust. and Appl. Math.*, vol. 3, pp. 192-253; December, 1955. See sec. 3. This paper contains a good bibliography. Also *J. Soc. Indust. and Appl. Math.*, vol. 4, pp. 86-119; June, 1956.

<sup>8</sup> D. Van Meter and D. Middleton, "Modern statistical approaches to reception in communication theory," *IRE TRANS.*, PGIT-4, pp. 19-145; September, 1954.

<sup>9</sup> L. A. Zadeh and J. R. Ragazzini, "Optimum filters for the detection of signals in noise," *PROC. IRE*, vol. 40, pp. 1223-1231; October, 1952.

assuming that the Gaussian channel noise has a "white" spectrum of unlimited extent. The Gaussian signals also assume a special form, as will be seen in the first part of the next section.

Most of the analysis to follow was originally pursued with the intention of finding an optimum detector for the transmission of members of a finite set of known band-pass waveforms over a channel containing multiple "scatter-path" modes of propagation and additive white Gaussian noise.<sup>10</sup> Since such scatter-modes perturb this transmission into a Gaussian signal, the same analysis serves both the scatter-multipath and random-signal detection problems, with the latter receiving the greater emphasis in this paper. Although the following work was thus done independently and in a different context from that of Davis, and employs sampling-point methods that may be somewhat less rigorous than the eigenfunction expansions used by Grenander and Davis, many of the results can be shown to agree with Davis' general expressions.

This paper is divided into two parts. Part I presents general and special derivations of the optimum, probability-computing detector (nondecision) from both the random-signal and scatter-multipath aspects. Part II<sup>11</sup> will give results on the performance of the Ideal decision-making detector in some special cases, together with comparisons of the performances of other decision detectors.

#### DETERMINATION OF THE OPTIMUM, PROBABILITY-COMPUTING DETECTOR FOR RANDOM TRANSMITTED SIGNALS AND FOR KNOWN SIGNALS TRANSMITTED THROUGH SCATTER MULTIPATH

##### *Form of Random Transmitted Signal and Its Relation to Scatter Transmission*

We assume throughout this paper that the random transmitted signal  $z^{(k)}(t)$  is given by

$$z^{(k)}(t) = \sum_{p=1}^N r^{(k)}(t - \tau_p) [y_p(t) * \theta^{(k)}(t - \tau_p)] \quad (1)$$

where  $[y_p(t) * \theta^{(k)}(t - \tau_p)]$  denotes phase modulation of a stationary narrow-band Gaussian process  $y_p(t)$  by  $\theta^{(k)}(t - \tau_p)$ :

$$y_p(t) * \theta^{(k)}(t - \tau_p) = y_{sp}(t) \sin [\omega_0 t + \theta^{(k)}(t - \tau_p)] + y_{cp}(t) \cos [\omega_0 t + \theta^{(k)}(t - \tau_p)] \quad (2)$$

using Rice's expression<sup>12</sup> for narrow-band Gaussian noise. Assuming for the remainder of this paper that the spectrum  $Y_p(\omega)$  of  $y_p(t)$  is symmetric about  $\omega_0$ ,  $y_{sp}(t)$  and  $y_{cp}(t)$  are independent low-pass Gaussian processes of zero mean and autocorrelation function  $\phi_p(\tau)$ :

$$\begin{aligned} \phi_p(\tau) &= \overline{y_{sp}(t)y_{sp}(t + \tau)} = \overline{y_{cp}(t)y_{cp}(t + \tau)} \\ &= \int_0^\infty Y_p(\omega) \cos(\omega - \omega_0)\tau d\omega. \end{aligned} \quad (3)$$

The amplitude and phase modulations,  $r^{(k)}(t)$  and  $\theta^{(k)}(t)$ , respectively, are members of a finite set of  $2M$  waveforms known *a priori* to the receiver. In conjunction with the  $y_p(t)$ , these modulations form the  $M$  ensemble from which  $z^{(k)}(t)$  is drawn, as given in the definition of the optimum detector.<sup>9</sup> The  $\tau_p$  are arbitrary delay constants.

Eq. (1) has been chosen for the transmitted signal in order to allow the results of this analysis to extend to the scatter-multipath detection problem. While (1) is consequently rather restricted in form, it covers a sufficiently broad class of random signals to be of interest.

To demonstrate the appropriateness of (1) to the scatter-multipath problem, let us suppose that a nonrandom, narrow-band signal  $x^{(k)}(t)$  is transmitted over scatter multipath

$$x^{(k)}(t) = r^{(k)}(t) \sin [\omega_0 t + \theta^{(k)}(t)]. \quad (4)$$

Assuming that the  $N$ -scatter paths have various delays,  $\tau_p$ , and that the path fluctuations are slow compared to  $\omega_0$ , present theory<sup>13</sup> indicates that the response of the  $p$ th path to an unmodulated carrier  $x_0(t) = \sin \omega_0 t$  is the Gaussian process  $y_p(t)$  discussed in connection with (2) and (3). Thus when  $x^{(k)}(t)$  is transmitted we obtain (1) after scattering over all  $N$  paths. A diagram of the complete communication system, illustrating the relationship between the random-transmitted-signal system and the scatter-communication system, is shown in Fig. 1, opposite.

##### *General Solution for the Optimum Detector*<sup>14</sup>

The received signal  $w(t) = z^{(k)}(t) + n(t)$ , where  $z^{(k)}(t)$  is given by (1) and  $n(t)$  is a white Gaussian noise of spectral density  $N_0/2\pi$  defined on the basis of a single-sided spectrum and an angular-frequency variable  $\omega$ .

<sup>10</sup> Earlier developments of the work contained in this paper may be found in:

R. Price, "Statistical Theory Applied to Communication Through Multipath Disturbances," M.I.T. Res. Lab. of Electronics, Tech. Rep. 266; September 3, 1953.

R. Price, "The detection of signals perturbed by scatter and noise," IRE TRANS., PGIT-4, pp. 163-170; September, 1954.

R. Price, M.I.T. Lincoln Lab. Group Reps. (Not generally available.)

Related studies have been carried on by:

W. L. Root and T. S. Pitcher, "Some remarks on statistical detection," IRE TRANS., vol. IT-1, pp. 33-38; December, 1955.

G. Turin, M.I.T. Lincoln Lab. Tech. Rep. (Not generally available. These results are to be published in IRE TRANS., PGIT.)

<sup>11</sup> To be published later in IRE TRANS. PGIT.

<sup>12</sup> S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282-332; July, 1944, and vol. 24, pp. 46-156; January, 1945. See sec. 3.7.

<sup>13</sup> S. O. Rice, "Statistical fluctuations of radio field strength far beyond the horizon," *Proc. IRE*, vol. 41, pp. 274-281; February, 1953.

<sup>14</sup> H. Sherman, "The Detection and Phase Determination of Signals in Additive and Multiplicative Noise," dissertation submitted in partial fulfillment of requirements for degree of doctor of electrical engineering, Brooklyn Polytech. Inst., p. 63; June, 1955. H. Sherman has carried out studies on the sampling point detection of signals in correlated Gaussian additive and multiplicative noise that are the discrete analog of this analysis. As in the case of Davis work, the results are largely in open form, since they involve the inversions of high-order matrices.

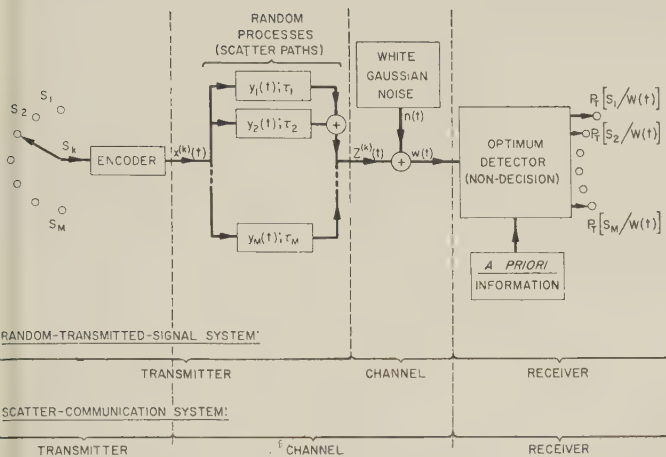


Fig. 1—The random-transmitted signal and scatter-communication systems.

We shall first assume that the noise is confined to a rectangular band of width  $2\pi B$  centered on  $\omega_0$ , and later allow  $B$  to increase without limit (ignoring spreading into the negative-frequency region), so that the assumption of white Gaussian noise is ultimately satisfied.

All the information in the band-pass signal  $w(t)$  is contained in its sine and cosine components,  $w_s(t)$  and  $w_c(t)$ , respectively. Sampling these components on a common-time grid of interval  $1/B$ , we obtain, using (1),

$$\left. \begin{aligned} w_{si} &= \sum_{p=1}^N x_{s,i-d_p}^{(k)} y_{sp} - \sum_{p=1}^N x_{c,i-d_p}^{(k)} y_{cp} + n_{si} \\ w_{ci} &= \sum_{p=1}^N x_{s,i-d_p}^{(k)} y_{cp} + \sum_{p=1}^N x_{c,i-d_p}^{(k)} y_{sp} + n_{ci} \end{aligned} \right\} \quad (5)$$

$$p_T[(w_{si}), (w_{ci})/(x_{si}^{(k)}), (x_{ci}^{(k)})]$$

$$= (2\pi)^{-n/2} |M^{(k)}|^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[ \frac{M_{ssij}^{(k)} w_{si} w_{sj} + M_{scij}^{(k)} w_{si} w_{cj} + M_{csij}^{(k)} w_{ci} w_{sj} + M_{ccij}^{(k)} w_{ci} w_{cj}}{|M^{(k)}|} \right] \right\} \quad (10)$$

where the subscript  $i$  denotes a sample value of the corresponding waveform at  $t = (i/B) + t_0$ , with the observation interval commencing at  $t_0$ . The functions

$$\left. \begin{aligned} x_s^{(k)}(t) &= r^{(k)}(t) \cos \theta^{(k)}(t) \\ x_c^{(k)}(t) &= r^{(k)}(t) \sin \theta^{(k)}(t) \end{aligned} \right\} \quad (6)$$

are the sine and cosine components of  $x^{(k)}(t)$ , respectively. Likewise,  $n_s(t)$  and  $n_c(t)$  are the sine and cosine components of  $n(t)$ , respectively. We may assume that the  $\tau_p$  are integral multiples of  $1/B$ , and we have defined

$$d_p \equiv \tau_p B. \quad (7)$$

Assuming that the  $y_p(t)$  processes (or paths) fluctuate independently of each other and of the noise, we find the second moments of  $w_{si}$  and  $w_{ci}$  to be, for  $s_k$  encoded at the transmitter,

$$\left. \begin{aligned} m_{ssij}^{(k)} &= \overline{w_{si} w_{sj}} \\ m_{ccij}^{(k)} &= \overline{w_{ci} w_{cj}} \end{aligned} \right\}$$

$$\left. \begin{aligned} &= \sum_{p=1}^N [x_{s,i-d_p}^{(k)} x_{s,i-d_p}^{(k)} + x_{c,i-d_p}^{(k)} x_{c,i-d_p}^{(k)}] \phi_{pii} + N_0 B \delta_{ii} \\ m_{ssij} &= \overline{w_{si} w_{sj}} \\ -m_{ccij} &= -\overline{w_{ci} w_{sj}} \end{aligned} \right\} \quad (8)$$

$$= \sum_{p=1}^N [x_{s,i-d_p}^{(k)} x_{c,i-d_p}^{(k)} - x_{c,i-d_p}^{(k)} x_{s,i-d_p}^{(k)}] \phi_{pii}$$

where  $\phi_{pii} = \phi_p[(i-j)/B]$  and  $\delta_{ii}$  is the Kronecker  $\delta$  function.

The inverse, or *a posteriori*, probabilities  $P_T[s_k/w(t)]$  are found from analysis carried out in the opposite, or "forward," direction; *i.e.*, assuming the cause  $s_k$  and studying the effect  $w(t)$ . Using Bayes' Identity,

$$P_T[s_k/w(t)] = \lim_{B \rightarrow \infty} \frac{P_k p_T[(w_{si}), (w_{ci})/(x_{si}^{(k)}), (x_{ci}^{(k)})]}{\sum_{r=1}^M P_r p_T[(w_{si}), (w_{ci})/(x_{si}^{(r)}), (x_{ci}^{(r)})]} \quad (9)$$

(infinitely dense sampling)

where parentheses denote the sequence of all sample values they enclose. The subscript  $T$  on the "forward," conditional, multivariate probability density  $p_T[(w_{si}), (w_{ci})/(x_{si}^{(k)}), (x_{ci}^{(k)})]$  indicates restriction of the  $(w_{si})$  and  $(w_{ci})$  sequences to sample values contained in the observation interval. In order to obtain the "forward" probability densities, we note, from (5), that since  $y_{sp}(t)$ ,  $y_{cp}(t)$ ,  $n_s(t)$  and  $n_c(t)$  are independent Gaussian processes, the sequences  $(w_{si})$  and  $(w_{ci})$  share a joint Gaussian distribution,<sup>15</sup> conditional upon  $x^{(k)}(t)$  being known. Thus, from Rice,<sup>16</sup>

where  $M^{(k)}$  is the symmetric moment matrix

$$M^{(k)} = \begin{bmatrix} m_{ss11}^{(k)} & m_{ss12}^{(k)} & \cdots & m_{ss1n}^{(k)} & m_{sc11}^{(k)} & m_{sc12}^{(k)} & \cdots & m_{sc1n}^{(k)} \\ m_{ss21}^{(k)} & m_{ss22}^{(k)} & \cdots & m_{ss2n}^{(k)} & m_{sc21}^{(k)} & m_{sc22}^{(k)} & \cdots & m_{sc2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{ssn1}^{(k)} & m_{ssn2}^{(k)} & \cdots & m_{ssnn}^{(k)} & m_{scn1}^{(k)} & m_{scn2}^{(k)} & \cdots & m_{scnn}^{(k)} \\ m_{cs11}^{(k)} & m_{cs12}^{(k)} & \cdots & m_{cs1n}^{(k)} & m_{cc11}^{(k)} & m_{cc12}^{(k)} & \cdots & m_{cc1n}^{(k)} \\ m_{cs21}^{(k)} & m_{cs22}^{(k)} & \cdots & m_{cs2n}^{(k)} & m_{cc21}^{(k)} & m_{cc22}^{(k)} & \cdots & m_{cc2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{csn1}^{(k)} & m_{csn2}^{(k)} & \cdots & m_{csnn}^{(k)} & m_{ccn1}^{(k)} & m_{ccn2}^{(k)} & \cdots & m_{ccnn}^{(k)} \end{bmatrix}. \quad (11)$$

<sup>15</sup> H. Cramer, "Mathematical Methods of Statistics," Princeton Univ. Press, Princeton, N. J., p. 313; 1951.

<sup>16</sup> S. O. Rice, "Mathematical analysis of random noise," *op. cit.*, sec. 2.9.

$M_{ssij}^{(k)}$ ,  $M_{scij}^{(k)}$ ,  $M_{csij}^{(k)}$ , and  $M_{ccij}^{(k)}$  are the cofactors of  $m_{ssij}^{(k)}$ ,  $m_{scij}^{(k)}$ ,  $m_{scij}^{(k)}$ , and  $m_{ccij}^{(k)}$ , respectively, in  $M^{(k)}$  and  $|M^{(k)}|$  is the determinant of  $M^{(k)}$ . There are  $n = BT$  sampling points in the observation interval of length  $T$ .

Substituting for the cofactors

$$\begin{aligned} h_{ssij}^{(k)} &= \delta_{ij} - N_0 B \frac{M_{ssij}^{(k)}}{|M^{(k)}|}; \\ h_{ccij}^{(k)} &= \delta_{ij} - N_0 B \frac{M_{ccij}^{(k)}}{|M^{(k)}|}; \\ h_{scij}^{(k)} &= -N_0 B \frac{M_{scij}^{(k)}}{|M^{(k)}|}; \\ h_{csij}^{(k)} &= -N_0 B \frac{M_{csij}^{(k)}}{|M^{(k)}|} \end{aligned} \quad (12)$$

we find, by straightforward matrix algebra, that  $h_{ssij}^{(k)}$  and  $h_{ccij}^{(k)}$  are also specified by the set of simultaneous equation pairs

$$\begin{cases} \sum_{i=1}^n [m_{ssli}^{(k)} h_{ssij}^{(k)} + m_{scli}^{(k)} h_{csij}^{(k)}] = m_{sslj}^{(k)} - N_0 B \delta_{lj} \\ \sum_{i=1}^n [m_{csli}^{(k)} h_{ssij}^{(k)} + m_{ccli}^{(k)} h_{csij}^{(k)}] = m_{cslj}^{(k)} \end{cases} \quad 1 \leq j, l \leq n \quad (13)$$

with  $h_{ccij}^{(k)}$  and  $h_{scij}^{(k)}$  given by a similar set. Using (8), it may be shown without difficulty, from these sets of equations,

$$\begin{cases} h_{ssij}^{(k)} = h_{ccij}^{(k)} = h_{ssji}^{(k)} = h_{ccji}^{(k)} \equiv h_{1ij}^{(k)} \\ h_{scij}^{(k)} = -h_{scji}^{(k)} = -h_{csij}^{(k)} = h_{csji}^{(k)} \equiv h_{2ij}^{(k)} \end{cases} \quad (14)$$

Substituting (10), (12), and (14) in (9), we find

$$P_T[s_k/w(t)] = \lim_{B \rightarrow \infty} \frac{\left[ \frac{|M^{(k)}|}{(N_0 B)^{2n}} \right]^{-1/2} P_k \exp \left\{ \frac{1}{2N_0 B} \sum_{i=1}^n \sum_{j=1}^n [(w_{si} w_{sj} + w_{ci} w_{cj}) h_{1ij}^{(k)} - (w_{si} w_{cj} - w_{ci} w_{sj}) h_{2ij}^{(k)}] \right\}}{\sum_{r=1}^M \left[ \frac{|M^{(r)}|}{(N_0 B)^{2n}} \right]^{-1/2} P_r \exp \left\{ \frac{1}{2N_0 B} \sum_{i=1}^n \sum_{j=1}^n [(w_{si} w_{sj} + w_{ci} w_{cj}) h_{1ij}^{(r)} - (w_{si} w_{cj} - w_{ci} w_{sj}) h_{2ij}^{(r)}] \right\}} \quad (15)$$

In order to find the limiting behavior of the  $|M^{(k)}|/(N_0 B)^{2n}$ , we note that, from the theory of determinants,

$$|M^{(k)}| = \prod_{l=1}^{2n} \bar{\lambda}_{nl}^{(k)} \quad (16)$$

where the  $\bar{\lambda}_{nl}^{(k)}$  are the eigenvalues of the homogeneous system of equations corresponding to (11):

$$\begin{cases} \sum_{j=1}^n [m_{ssij}^{(k)} \psi_{1lj}^{(k)} + m_{scij}^{(k)} \psi_{2lj}^{(k)}] = \bar{\lambda}_{nl}^{(k)} \psi_{1li}^{(k)} \\ \sum_{j=1}^n [m_{csij}^{(k)} \psi_{1lj}^{(k)} + m_{ccij}^{(k)} \psi_{2lj}^{(k)}] = \bar{\lambda}_{nl}^{(k)} \psi_{2li}^{(k)} \end{cases} \quad i = 1, 2, \dots, n \quad (17)$$

and the pair of sequences  $\psi_{1li}^{(k)}$ ,  $\psi_{2li}^{(k)}$ , with  $l = 1, 2, \dots, 2n$ , is the eigenvector corresponding to the eigenvalue  $\bar{\lambda}_{nl}^{(k)}$ .

Letting

$$\bar{\lambda}_{nl}^{(k)} = \lambda_{nl}^{(k)} + N_0 B \quad (18)$$

we have

$$\frac{|M^{(k)}|}{(N_0 B)^{2n}} = \prod_{l=1}^{2n} \left( 1 + \frac{\lambda_{nl}^{(k)}}{N_0 B} \right) \quad (19)$$

with (17) modified appropriately for the new  $\lambda_{nl}^{(k)}$ .

Carrying out the limiting process of (15), the summations of (13), (15), and (17) become integrals, with limits  $t_0$  and  $t_1 = t_0 + T$ , and we have

$$P_T[s_k/w(t)] = \frac{P_k L_k}{\sum_{r=1}^M P_r L_r} \quad (20)$$

where

$$\begin{aligned} L_k &= F_k^{-1/2} \exp \left[ \frac{1}{2N_0} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \{ [w_s(t) w_s(\tau) \right. \\ &\quad + w_c(t) w_c(\tau)] h_1^{(k)}(t - \tau, t) \\ &\quad - [w_s(t) w_c(\tau) - w_c(t) w_s(\tau)] h_2^{(k)}(t - \tau, t) \} d\tau dt \left. \right] \quad (21) \end{aligned}$$

$h_1^{(k)}(t - \tau, t)$  and  $h_2^{(k)}(t - \tau, t)$  satisfy, from (8) and (13)

$$\begin{cases} \int_{t_0}^{t_1} \{ [\Phi_1^{(k)}(\sigma, \tau) + N_0 \delta(\sigma - \tau)] h_1^{(k)}(t - \tau, t) \\ - \Phi_2^{(k)}(\sigma, \tau) h_2^{(k)}(t - \tau, t) \} d\tau = \Phi_1^{(k)}(\sigma, t) \\ t_0 \leq \sigma, t \leq t_1 \end{cases} \quad (22)$$

where  $\delta(t)$  is the Dirac  $\delta$  function. We have made the substitutions

$$\begin{cases} \Phi_1^{(k)}(\sigma, \tau) = \sum_{p=1}^N X_{1pp}^{(k)} \phi_p(\sigma - \tau) \\ \Phi_2^{(k)}(\sigma, \tau) = \sum_{p=1}^N X_{2pp}^{(k)} \phi_p(\sigma - \tau) \end{cases} \quad (23)$$

and

$$\begin{cases} X_{1pq}^{(k)}(\sigma, \tau) = x_s^{(k)}(\sigma - \tau_p) x_s^{(k)}(\tau - \tau_q) \\ \quad + x_c^{(k)}(\sigma - \tau_p) x_c^{(k)}(\tau - \tau_q) \\ X_{2pq}^{(k)}(\sigma, \tau) = x_s^{(k)}(\sigma - \tau_p) x_c^{(k)}(\tau - \tau_q) \\ \quad - x_c^{(k)}(\sigma - \tau_p) x_s^{(k)}(\tau - \tau_q) \end{cases} \quad (24)$$

From (17) and (19),

$$F_k = \prod_{l=1}^{\infty} \left( 1 + \frac{\lambda_l^{(k)}}{N_0} \right) \quad (25)$$

where the  $\lambda_l^{(k)}$  are the eigenvalues of the pair of integral equations,

$$\left. \begin{aligned} & [\Phi_1^{(k)}(\sigma, \tau) \psi_{1l}^{(k)}(\tau) + \Phi_2^{(k)}(\sigma, \tau) \psi_{2l}^{(k)}(\tau)] d\tau \\ & = \lambda_l^{(k)} \psi_{1l}^{(k)}(\sigma) \\ & t_0 \leq \sigma \leq t_1 \\ & [-\Phi_2^{(k)}(\sigma, \tau) \psi_{1l}^{(k)}(\tau) + \Phi_1^{(k)}(\sigma, \tau) \psi_{2l}^{(k)}(\tau)] d\tau \\ & = \lambda_l^{(k)} \psi_{2l}^{(k)}(\sigma) \end{aligned} \right\} \quad (26)$$

and the pair of functions  $\psi_{1l}^{(k)}, \psi_{2l}^{(k)}$  gives the eigenfunction corresponding to  $\lambda_l^{(k)}$ .

Eqs. (20) through (26) constitute the solution for the optimum probability-computing detector in the general case of multiprocess random transmitted signals, or for scatter multipath. The simultaneous integral (22) and (26) can be solved in special cases to be discussed in the remainder of Part I.

We may note three interesting results of a general nature that are obtainable from (22). First, if either

$$\left. \begin{aligned} x_s^{(k)}(t - \tau_p) &= 0 = x_c^{(k)}(t - \tau_p) \\ x_s^{(k)}(\tau - \tau_p) &= 0 = x_c^{(k)}(\tau - \tau_p) \end{aligned} \right\} \text{ for all } p \quad (27)$$

or both,  $h_1^{(k)}(t - \tau, t)$  and  $h_2^{(k)}(t - \tau, t)$  vanish. Thus, if the waveform  $x^{(k)}(t)$  vanishes over an interval long enough that only the white Gaussian noise is received during a certain period, the receiver may ignore that period with no loss of information about  $s_k$ . This result agrees with intuition.

A second and more subtle result is that the double integration of (21) can be performed by physically-realizable filters, in the sense that  $h_1^{(k)}(t - \tau, t)$  and  $h_2^{(k)}(t - \tau, t)$  may be allowed to vanish for  $\tau > t$ . For, from (24),

$$\left. \begin{aligned} h_1^{(k)}(t - \tau, t) &= h_1^{(k)}(\tau - t, \tau) \\ h_2^{(k)}(t - \tau, t) &= -h_2^{(k)}(\tau - t, \tau) \end{aligned} \right\} \quad (28)$$

Thus, the integrand of (21) is symmetric in  $t$  and  $\tau$ . Since for any symmetric function  $K(t, \tau) = K(\tau, t)$ ,

$$\int_{t_0}^{t_1} \int_{t_0}^{t_1} K(t, \tau) d\tau dt = 2 \int_{t_0}^{t_1} \int_{t_0}^t K(t, \tau) d\tau dt \quad (29)$$

and the upper limit on the inner integral in (21) may be changed from  $t_1$  to  $t$  [provided that the factor of 2 in (29) is allowed for], thus permitting  $h_1^{(k)}(t - \tau, t)$  and  $h_2^{(k)}(t - \tau, t)$  to vanish for  $\tau > t$ .

Finally, it may easily be shown that the form of (21) is invariant to a purely rotational transformation of  $w_s(t)$  and  $w_c(t)$ , a result justified intuitively by the complete randomness of initial phase in  $z^{(k)}(t)$ .

#### Special Cases of Multiprocess Signals (or Scatter Multipath)

**Amplitude-Modulated  $x^{(k)}(t)$ :** If  $x^{(k)}(t)$  is a purely amplitude-modulated carrier,  $\Phi_2^{(k)}(\sigma, \tau)$  vanishes, so that (25) and (26) become, respectively,

$$\left. \begin{aligned} & \int_{t_0}^{t_1} \left[ \sum_{p=1}^N r^{(k)}(\sigma - \tau_p) r^{(k)}(\tau - \tau_p) \phi_p(\sigma - \tau) \right. \\ & \quad \left. + N_0 \delta(\sigma - \tau) \right] h_1^{(k)}(t - \tau, t) d\tau \\ & = \sum_{p=1}^N [r^{(k)}(\sigma - \tau_p) r^{(k)}(t - \tau_p) \phi_p(\sigma - t)] \\ & \quad t_0 \leq \sigma, t \leq t_1 \\ & h_2^{(k)}(t - \tau, t) = 0 \end{aligned} \right\} \quad (30)$$

$$\int_{t_0}^{t_1} \left[ \sum_{p=1}^N r^{(k)}(\sigma - \tau_p) r^{(k)}(\tau - \tau_p) \phi_p(\sigma - \tau) \right] \psi_l^{(k)}(\tau) d\tau = \lambda_l^{(k)} \psi_l^{(k)}(\sigma); \quad t_0 \leq \sigma \leq t_1 \quad (31)$$

with each  $\lambda_l^{(k)}$  now counted twice in (25). If, in addition,  $r^{(k)}(t - \tau_p) = \gamma$ , a constant, for all  $t_0 \leq t \leq t_1$ , and for all  $p$ , and if the  $\phi_p(\tau)$  all have rational algebraic Fourier transforms  $Y_p(\omega)$ , (30) and (31) can in principle be solved by the methods of Zadeh and Ragazzini,<sup>8,17</sup> Middleton,<sup>18</sup> Youla,<sup>19,20</sup> Slepian,<sup>21</sup> and Miller and Zadeh.<sup>22</sup>

**Very Slowly Changing Processes or Scatter:** If all the  $y_p(t)$  processes have such narrow spectra  $Y_p(\omega)$  that they are nearly sinusoidal during the observation interval, with negligible amplitude or phase modulation, we may set  $\phi_p(\tau) = \phi_p$ , a constant, for  $|\tau| \leq T$ . Eq. (22) may then be solved by letting

$$\left. \begin{aligned} h_1^{(k)}(t - \tau, t) &= \sum_{p=1}^N \sum_{q=1}^N [a_{pq}^{(k)} X_{1pq}^{(k)}(\tau, t) + b_{pq}^{(k)} X_{2pq}^{(k)}(\tau, t)] \\ h_2^{(k)}(t - \tau, t) &= \sum_{p=1}^N \sum_{q=1}^N [a_{pq}^{(k)} X_{2pq}^{(k)}(\tau, t) - b_{pq}^{(k)} X_{1pq}^{(k)}(\tau, t)] \end{aligned} \right\} \quad (32)$$

Substituting (32) in (22), we find that both integral equations of (22) are satisfied if

$$\left. \begin{aligned} \sum_{p=1}^N [a_{pq}^{(k)} D_{pr}^{(k)} + b_{pq}^{(k)} E_{pr}^{(k)}] &= \delta_{rq}/N_0 \\ \sum_{p=1}^N [-a_{pq}^{(k)} E_{pr}^{(k)} + b_{pq}^{(k)} D_{pr}^{(k)}] &= 0 \end{aligned} \right\} 1 \leq q, r \leq N \quad (33)$$

where

$$\left. \begin{aligned} D_{pr}^{(k)} &= \frac{1}{N_0} \int_{t_0}^{t_1} X_{1pr}^{(k)}(t, t) dt + \delta_{pr}/\phi_r \\ E_{pr}^{(k)} &= \frac{1}{N_0} \int_{t_0}^{t_1} X_{2pr}^{(k)}(t, t) dt \end{aligned} \right\} \quad (34)$$

<sup>17</sup> L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645-655; July, 1950.

<sup>18</sup> D. Middleton, M.I.T. Lincoln Lab. Tech. Rep. (Not generally available.)

<sup>19</sup> D. Youla, "The use of the method of maximum likelihood in estimating continuous-modulated intelligence which has been corrupted by noise," *IRE TRANS., PGIT-3*, pp. 90-105; March, 1954.

<sup>20</sup> D. Youla, "A Finite-Time Homogeneous Wiener-Hopf Integral Equation," Res. Rep. R-445-55, PIB-376, Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., October 3, 1955.

<sup>21</sup> D. Slepian, "Estimation of signal parameters in the presence of noise," *IRE TRANS., PGIT-3*, pp. 68-89; March, 1954.

<sup>22</sup> K. S. Miller and L. A. Zadeh, "Solution of an integral equation occurring in the theories of prediction and detection," *IRE Trans.*, vol. IT-2, pp. 72-75; June, 1956.

Solving (33) for the  $a_{pq}^{(k)}$ ,

$$a_{pq}^{(k)} = \frac{M_{pq}^{(k)}}{N_0 |M^{(k)}|} \quad (35)$$

where  $M^{(k)}$  is the symmetric matrix

$$M^{(k)} = \begin{bmatrix} [D_{pq}^{(k)}] & [E_{pq}^{(k)}] \\ [-E_{pq}^{(k)}] & [D_{pq}^{(k)}] \end{bmatrix}. \quad (36)$$

$|M^{(k)}|$  is the determinant of  $M^{(k)}$  and  $M_{pq}^{(k)}$  is the cofactor of  $D_{pq}^{(k)}$  in (36). Substitution of (32) in (21) yields terms involving the  $a_{pq}^{(k)}$  only, so that it is not necessary to find the  $b_{pq}^{(k)}$  explicitly.

To find the  $F_k$  of (21), we let

$$\left. \begin{aligned} \psi_{1l}^{(k)}(\tau) &= \sum_{q=1}^N [A_{lq}^{(k)} x_s^{(k)}(\tau - \tau_q) + B_{lq}^{(k)} x_c^{(k)}(\tau - \tau_q)] \\ \psi_{2l}^{(k)}(\tau) &= \sum_{q=1}^N [-B_{lq}^{(k)} x_s^{(k)}(\tau - \tau_q) + A_{lq}^{(k)} x_c^{(k)}(\tau - \tau_q)] \end{aligned} \right\}. \quad (37)$$

Both integral equations of (26) are then satisfied if  $A_{lq}^{(k)}$  and  $B_{lq}^{(k)}$  satisfy

$$\left. \begin{aligned} \phi_p \sum_{q=1}^N [A_{lq}^{(k)} D_{pq}^{(k)} + B_{lq}^{(k)} E_{pq}^{(k)}] \\ &= [1 + \lambda_l^{(k)}/N_0] A_{lp}^{(k)} \\ \phi_p \sum_{q=1}^N [-B_{lq}^{(k)} E_{pq}^{(k)} + A_{lq}^{(k)} D_{pq}^{(k)}] \\ &= [1 + \lambda_l^{(k)}/N_0] B_{lp}^{(k)} \end{aligned} \right\} 1 \leq p \leq N. \quad (38)$$

Since the quantities  $[1 + \lambda_l^{(k)}/N_0]$  are the eigenvalues of the system of (38), and there are therefore  $2N$  of them, their product is the value of the determinant corresponding to the system (38). Thus

$$F_k = \prod_{l=1}^{2N} [1 + \lambda_l^{(k)}/N_0] = \left[ \prod_{p=1}^N \phi_p^2 \right] |M^{(k)}| \quad (39)$$

since the systems (33) and (38) are similar. Substituting in (32), (35), and (39) in (21), we have, finally,

$$L_k = \left[ \prod_{p=1}^N \phi_p \right]^{-1} |M^{(k)}|^{-1/2} \cdot \exp \left\{ \frac{1}{2} \sum_{p=1}^N \sum_{q=1}^N \frac{M_{pq}^{(k)}}{|M^{(k)}|} [A_p^{(k)} A_q^{(k)} + B_p^{(k)} B_q^{(k)}] \right\} \quad (40)$$

where

$$\left. \begin{aligned} A_p^{(k)} &= \frac{1}{N_0} \int_{t_0}^{t_1} W_{1p}^{(k)}(t) dt \\ B_p^{(k)} &= \frac{1}{N_0} \int_{t_0}^{t_1} W_{2p}^{(k)}(t) dt \end{aligned} \right\} \quad (41)$$

and

$$\left. \begin{aligned} W_{1p}^{(k)}(t) &= x_s^{(k)}(t - \tau_p) w_s(t) + x_c^{(k)}(t - \tau_p) w_c(t) \\ W_{2p}^{(k)}(t) &= x_s^{(k)}(t - \tau_p) w_c(t) - x_c^{(k)}(t - \tau_p) w_s(t) \end{aligned} \right\}. \quad (42)$$

This result appears in a different form in some recent work of Turin's.<sup>10</sup>

*Infinite Series Solution for Large  $N_0$ :* For sufficiently large  $N_0$ , (22) and (25) may be solved by developing

infinite series. By successive approximations we find from (22),

$$\begin{aligned} h_1^{(k)}(t - \tau, t) &= \frac{1}{N_0} \Phi_1^{(k)}(\tau, t) - \frac{1}{N_0^2} \\ &\cdot \int_{t_0}^{t_1} [\Phi_1^{(k)}(\tau, \sigma) \Phi_1^{(k)}(\sigma, t) - \Phi_2^{(k)}(\tau, \sigma) \Phi_2^{(k)}(\sigma, t)] d\sigma \\ &+ \text{terms of order } N_0^{-3} \text{ and smaller} \end{aligned}$$

$$\begin{aligned} h_2^{(k)}(t - \tau, t) &= \frac{1}{N_0} \Phi_2^{(k)}(\tau, t) - \frac{1}{N_0^2} \\ &\cdot \int_{t_0}^{t_1} [\Phi_1^{(k)}(\tau, \sigma) \Phi_2^{(k)}(\sigma, t) + \Phi_2^{(k)}(\tau, \sigma) \Phi_1^{(k)}(\sigma, t)] d\sigma \\ &+ \text{terms of order } N_0^{-3} \text{ and smaller.} \end{aligned} \quad (43)$$

$F_k$  may be approximated by

$$\begin{aligned} \log_e F_k &= \frac{1}{N_0} \sum_{l=1}^{\infty} \lambda_l^{(k)} - \frac{1}{2N_0^2} \sum_{l=1}^{\infty} [\lambda_l^{(k)}]^2 \\ &+ \text{terms of order } N_0^{-3} \text{ and smaller.} \end{aligned} \quad (44)$$

Examination of (26) shows that it is really a single homogeneous integral equation, despite the dual form. Hence sums of the form

$$\sum_{l=1}^{\infty} [\lambda_l^{(k)}]^j, \quad j = 1, 2, \dots,$$

may be found<sup>23</sup> by successive quadratures of the kernel of the single equation obtainable from (26). We find, for example,

$$\sum_{l=1}^{\infty} \lambda_l^{(k)} = 2 \sum_{p=1}^N \phi_p(0) \int_{t_0}^{t_1} [r^{(k)}(\sigma - \tau_p)]^2 d\sigma \quad (45)$$

$$\sum_{l=1}^{\infty} [\lambda_l^{(k)}]^2 = 2 \int_{t_0}^{t_1} \int_{t_0}^{t_1} \{ [\Phi_1^{(k)}(\sigma, \tau)]^2 + [\Phi_2^{(k)}(\sigma, \tau)]^2 \} d\tau d\sigma \quad (46)$$

Substituting (43) through (46) in (21),

$$\begin{aligned} \log_e L_k &\approx \frac{1}{2N_0^2} \sum_{p=1}^N \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left\{ W_{1p}^{(k)}(t) \left[ W_{1p}^{(k)}(\tau) \phi_p(t - \tau) \right. \right. \\ &- \frac{1}{N_0} \sum_{q=1}^N W_{1q}^{(k)}(\tau) U_{1pq}^{(k)}(t, \tau) \\ &+ \frac{1}{N_0} \sum_{q=1}^N W_{2q}^{(k)}(\tau) U_{2pq}^{(k)}(t, \tau) \left. \right] \\ &+ W_{2p}^{(k)}(t) \left[ W_{2p}^{(k)}(\tau) \phi_p(t - \tau) - \frac{1}{N_0} \sum_{q=1}^N W_{2q}^{(k)}(\tau) U_{1pq}^{(k)}(t, \tau) \right. \\ &- \frac{1}{N_0} \sum_{q=1}^N W_{1q}^{(k)}(\tau) U_{2pq}^{(k)}(t, \tau) \left. \right] \left. \right\} d\tau dt \\ &- \frac{1}{N_0} \sum_{p=1}^N \phi_p(0) \int_{t_0}^{t_1} [r^{(k)}(\sigma - \tau_p)]^2 d\sigma \\ &+ \frac{1}{2N_0^2} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \{ [\Phi_1^{(k)}(\sigma, \tau)]^2 + [\Phi_2^{(k)}(\sigma, \tau)]^2 \} d\tau d\sigma \end{aligned} \quad (47)$$

<sup>23</sup> R. Courant and D. Hilbert, "Methods of Mathematical Physics," Interscience Publishers, New York, N. Y., p. 138; 195 [Integrate (58) over  $s$  and  $t$ .]

to a second-order approximation, where

$$\left. \begin{aligned} U_{1pq}^{(k)}(t, \tau) &= \int_{t_0}^{t_1} X_{1pq}^{(k)}(\sigma, \sigma) \phi_p(\sigma - t) \phi_q(\sigma - \tau) d\sigma \\ U_{2pq}^{(k)}(t, \tau) &= \int_{t_0}^{t_1} X_{2pq}^{(k)}(\sigma, \sigma) \phi_p(\sigma - t) \phi_q(\sigma - \tau) d\sigma \end{aligned} \right\} \quad (48)$$

and  $W_{1p}^{(k)}(t)$  and  $W_{2p}^{(k)}(t)$  are given by (42). For very large  $N_0$  we have, from (47), the first-order approximation,

$$\begin{aligned} \log_e L_k \approx & \sum_{p=1}^N \left\{ \frac{1}{2N_0^2} \int_{t_0}^{t_1} \int_{t_0}^{t_1} [W_{1p}^{(k)}(t) W_{1p}^{(k)}(\tau) \right. \\ & + W_{2p}^{(k)}(t) W_{2p}^{(k)}(\tau)] \phi_p(t - \tau) d\tau dt \\ & \left. - \frac{\phi_p(0)}{N_0} \int_{t_0}^{t_1} [r^{(k)}(\sigma - \tau_p)]^2 d\sigma \right\}. \end{aligned} \quad (49)$$

Thus, at large  $N_0$  the optimum detector analyzes the contribution of each path separately, as though it were the only path present, and it is only when  $N_0$  decreases that interaction terms such as those of order  $N_0^{-3}$  in (47) become significant.

In Fig. 2, one possible form is presented for an analog

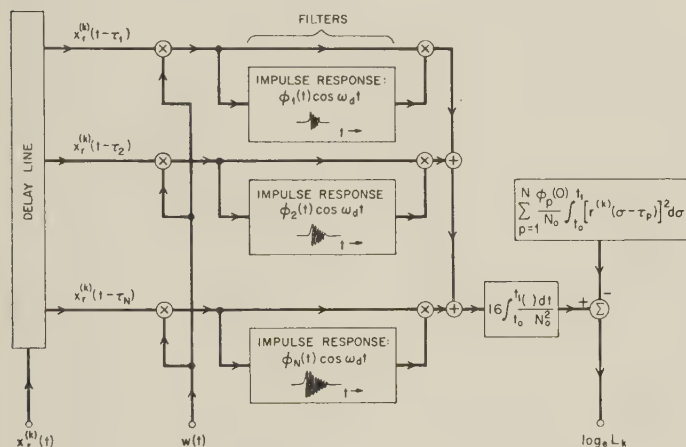


Fig. 2—First-order approximation to  $k$ th computing element of optimum receiver.

computer to calculate (49). We assume that the signal  $z^{(k)}(t)$  has, as before, the carrier frequency  $\omega_0$ , but that the stored reference waveform  $x_r^{(k)}(t)$  differs from the transmitter  $x^{(k)}(t)$  in having a carrier frequency  $\omega_r = \omega_0 \pm \omega_d$ ,

$$x_r^{(k)}(t) = x_s^{(k)}(t) \sin \omega_r t + x_c^{(k)}(t) \cos \omega_r t. \quad (50)$$

Assuming that  $\omega_d$  is small compared to  $\omega_0$  but large compared to  $1/T$  and to the highest-frequency components of  $z_s^{(k)}(t)$  and  $z_c^{(k)}(t)$ , it may be verified that

$$\begin{aligned} & \int_{t_0}^{t_1} x_r^{(k)}(t - \tau_p) w(t) dt \\ & \quad \cdot \int_{t_0}^t x_r^{(k)}(\tau - \tau_p) w(\tau) \phi_p(t - \tau) \cos \omega_d(t - \tau) d\tau \\ &= \frac{1}{32} \int_{t_0}^{t_1} \int_{t_0}^{t_1} [W_{1p}^{(k)}(t) W_{1p}^{(k)}(\tau) \\ & \quad + W_{2p}^{(k)}(t) W_{2p}^{(k)}(\tau)] \phi_p(t - \tau) d\tau dt \end{aligned} \quad (51)$$

with the substitution of the variable limit  $t$  as discussed in connection with (29).

### Optimum Detection for a Single-Process Random Signal (or a Single Scatter Path)

When there is only one random process or path present,  $N = 1$ , and (21) and (22) may be simplified by letting

$$\left. \begin{aligned} h_1^{(k)}(t - \tau, t) &= X_{111}^{(k)}(\tau, t) h^{(k)}(t - \tau, t) \\ h_2^{(k)}(t - \tau, t) &= X_{211}^{(k)}(\tau, t) h^{(k)}(t - \tau, t) \end{aligned} \right\}. \quad (52)$$

Eq. (21) then becomes

$$\begin{aligned} L_k = F_k^{-1/2} \exp \left\{ \frac{1}{2N_0} \int_{t_0}^{t_1} \int_{t_0}^{t_1} [W_{11}^{(k)}(t) W_{11}^{(k)}(\tau) \right. \\ \left. + W_{21}^{(k)}(t) W_{21}^{(k)}(\tau)] h^{(k)}(t - \tau, t) d\tau dt \right\} \end{aligned} \quad (53)$$

and both equations of (22) yield

$$\begin{aligned} & \int_{t_0}^{t_1} \{ [r^{(k)}(\tau - \tau_1)]^2 \phi_1(\sigma - \tau) \\ & \quad + N_0 \delta(\sigma - \tau) \} h^{(k)}(t - \tau, t) d\tau \\ &= \phi_1(\sigma - \tau); \quad t_0 \leq \sigma, t \leq t_1. \end{aligned} \quad (54)$$

Similarly, letting

$$\left. \begin{aligned} \psi_{1i}^{(k)}(\tau) &= x_s^{(k)}(\tau - \tau_1) \psi_i^{(k)}(\tau) \\ \psi_{2i}^{(k)}(\tau) &= x_c^{(k)}(\tau - \tau_1) \psi_i^{(k)}(\tau) \end{aligned} \right\} \quad (55)$$

both equations of (26) yield

$$\begin{aligned} & \int_{t_0}^{t_1} [r^{(k)}(\tau - \tau_1)]^2 \phi_1(\sigma - \tau) \psi_i^{(k)}(\tau) d\tau \\ &= \lambda_i^{(k)} \psi_i^{(k)}(\sigma); \quad t_0 \leq \sigma \leq t_1 \end{aligned} \quad (56)$$

and each of the  $\lambda_i^{(k)}$  eigenvalues of (56) should be counted twice in (25).

In general, the solutions of (54) and (56) are not known. For particular  $r^{(k)}(t)$  and  $\phi_1(\tau)$ , however, they may be solved, and these special cases will be considered in the next section.

We may, however, draw one very interesting general result from (53) and (54). If we assume that  $s_k$  is to be encoded at the transmitter, and in (42) substitute for  $w_s(t)$  and  $w_c(t)$  the expressions corresponding to the continuous form of (5), we obtain

$$\left. \begin{aligned} W_{11}^{(k)}(t)/r^{(k)}(t - \tau_1) &= r^{(k)}(t - \tau_1) y_{s1}(t) + n'_s(t) \\ W_{21}^{(k)}(t)/r^{(k)}(t - \tau_1) &= r^{(k)}(t - \tau_1) y_{c1}(t) + n'_c(t) \end{aligned} \right\} \quad (57)$$

where  $n'_s(t)$  and  $n'_c(t)$  are independent white Gaussian noises statistically identical to  $n_s(t)$  and  $n_c(t)$ . Now if, somehow,  $y_{s1}(t)$  and  $y_{c1}(t)$  were completely known to the receiver *a priori*, for  $t_0 \leq t \leq t_1$ , rather than just their statistics, the optimum detector would compute, according to a simple extension of the theory of Woodward and Davies,<sup>24</sup> for the single channel disturbed solely by white Gaussian noise,

<sup>24</sup> P. M. Woodward, "Probability and Information Theory, with Applications to Radar," McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.

$$P_T[s_k' w(t)] \Big|_{\substack{y_{s1}(t) \\ \text{known}}} = \frac{P_k L'_k}{\sum_{\tau=1}^M P_\tau L'_\tau} \quad (58)$$

where

$$L'_k = \exp \left\{ -\frac{1}{2N_0} \int_{t_0}^{t_1} [y_{s1}^2(t) + y_{e1}^2(t)] [r^{(k)}(t - \tau_1)]^2 dt \right\} \\ \cdot \exp \left\{ \frac{1}{N_0} \int_{t_0}^{t_1} [W_{11}^{(k)}(t) y_{s1}(t) + W_{21}^{(k)}(t) y_{e1}(t)] dt \right\}. \quad (59)$$

been solved explicitly for particular forms of  $r^{(k)}(t)$  and  $\phi_1(\tau)$ , and these solutions are now presented.<sup>26</sup>

Case I. Solution for Constant  $r^{(k)}(t)$  and Exponential  $\phi_1(\tau)$ : For

$$r^{(k)}(t - \tau_1) = \gamma, \quad \text{a constant} \quad (61)$$

$$\phi_1(\tau) = \beta e^{-\alpha|\tau|}. \quad (62)$$

Eq. (54) can be solved by the method of Zadeh and Ragazzini,<sup>17</sup> yielding

$$\frac{N_0 c}{\beta} h^{(k)}(t - \tau, t) = \frac{2 \left( \frac{c-1}{c+1} \right)^2 \exp(-2\alpha c T) \cosh[\alpha c(t - \tau)] + 2 \left( \frac{c-1}{c+1} \right) \exp(-\alpha c T) \cosh[\alpha c(t + \tau - t_0 - t_1)]}{1 - \left( \frac{c-1}{c+1} \right)^2 \exp(-2\alpha c T)} \\ + \exp[-\alpha c |t - \tau|] \quad (63)$$

Comparing the  $w(t)$ -dependent portions of (53) and (59), where we may draw the conjecture that the integrals

$$\left. \begin{aligned} I_1(t) &= \int_{t_0}^{t_1} W_{11}^{(k)}(\tau) h^{(k)}(t - \tau, t) d\tau \\ I_2(t) &= \int_{t_0}^{t_1} W_{21}^{(k)}(\tau) h^{(k)}(t - \tau, t) d\tau \end{aligned} \right\} \quad (60) \quad \text{and}$$

yield estimates of  $y_{s1}(t)$  and  $y_{s2}(t)$ , respectively, for the actual system considered. This conjecture proves to be correct, and, in fact, the integrals (60) are both maximum-likelihood and least-mean-square-error estimates.<sup>25</sup> Referring to Youla's work<sup>19</sup> on the maximum-likelihood estimation of a modulated Gaussian waveform in noise, we see that Youla's (1), (25), and (26) are identical to (57), (54), and (60) of this paper, respectively, with the following correspondences:

Youla	This Paper	Youla	This Paper
$M(t)$	$r^{(k)}(t - \tau_1)$	$W(t, \tau)$	$h^{(k)}(t - \tau, t) r^{(k)}(\tau - \tau_1)$
$a(t)$	$y_{s1}(t)$ or $y_{e1}(t)$	$t - T; t$	$t_0; t_1$
$n(t)$	$n_s'(t)$ or $n_e'(t)$	$a^*(t)$	$I_1(t)$ or $I_2(t)$
$e_1(t)$	$\begin{cases} W_{11}^{(k)}(t)/r^{(k)}(t - \tau_1) \\ \text{or } W_{21}^{(k)}(t)/r^{(k)}(t - \tau_1) \end{cases}$	$R_a(t, \tau)$	$\phi_1(t - \tau)$
		$R_n(t, \tau)$	$N_0 \delta(t - \tau)$

The same general result has also been obtained<sup>10</sup> using sampling-point analysis rather than the eigenfunction expansions employed by Youla.

#### Special Cases of Complete Solution for Single Process or Scatter Path

Under "Special Cases of Multiprocess Signals (or Scatter Multipath)," we discussed three special cases in which the multiprocess or multipath integral equations (22) and (26) can be solved, at least in principle. These solutions exist also, of course, for the single-process or single-path equations (54) and (56). In addition, (54) and (56) have

<sup>25</sup> The suggestion of using a correlator in conjunction with an estimator for a multipath receiver first appeared in studies whose results are presented in Root and Pitcher, *op. cit.*

$$c = \sqrt{1 + \frac{2}{\eta}} \quad (64)$$

$$\eta = N_0 \alpha / (\beta \gamma^2). \quad (65)$$

Youla<sup>19</sup> has given (63) in his (30), in slightly different notation and for  $\gamma = 1$ . The dependence of the shape of  $h^{(k)}(t - \tau, t)$  on  $\eta$  is illustrated in Fig. 3 (opposite).

The solution of (56) for  $r^{(k)}(t)$  and  $\phi_1(\tau)$  given by (61) and (62), respectively, has been found by Kac and Siegert<sup>27</sup> and by Slepian.<sup>21</sup> The eigenvalues are

$$\lambda_l^{(k)} = \frac{2\beta\gamma^2 T}{8(1 + R_l^2)} \quad (66)$$

where the  $R_l$  are the positive roots of the transcendental equation

<sup>26</sup> A. J. F. Siegert, "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena. II. Examples," Rep. P-730, The RAND Corp., Santa Monica, Calif., September, 1955. "Passage of stationary processes through linear and non-linear devices," IRE Trans., vol. PGIT-3, pp. 4-25; March, 1954.

D. Middleton has called the author's attention to these two pertinent papers. Herein, closed forms for products such as (25.) are obtained without having to solve (56.). Thus for Case I we find, from (2.32) of the RAND report,

$$F_K^{1/2} = e^{-g} (\cosh gc + e^{-1} \left[ 1 + \frac{1}{g\bar{N}_0} \right] \sinh gc)$$

with  $c$  and  $\bar{N}_0$  given by (64.) and (70.), respectively. For Case II, from (2.48) of the RAND report,

$$F_K^{1/2} = \frac{2}{\bar{N}_0} (\kappa)^{1+\nu} [I_{\nu-1}(d) K_{\nu+1}(\kappa d) - K_{\nu-1}(d) I_{\nu+1}(\kappa d)]$$

with  $\kappa = \exp(-\mu T)$  and  $d = \sqrt{4\nu/\bar{N}_0}$ . The  $I$ 's and  $K$ 's are modified Bessel functions and  $\nu$  and  $\bar{N}_0$  are given by (73.) and (82.), respectively. In both cases the  $\lambda^{(k)}_l$  are counted twice in (25.), as stated following (56.).

In general, it appears that since  $h^{(k)}(t - \tau, t)$  is the Volterra Reciprocal Function yielding the Fredholm Determinant  $F^{1/2}_{k_1}$ , a complete solution of the single-process or single-path case need deal only with (54.), rather than both (54.) and (56.).

<sup>27</sup> M. Kac and A. J. F. Siegert, "On the theory of noise in radio receivers with square law detectors," J. Appl. Phys., vol. 18, pp. 383-397; April, 1947.

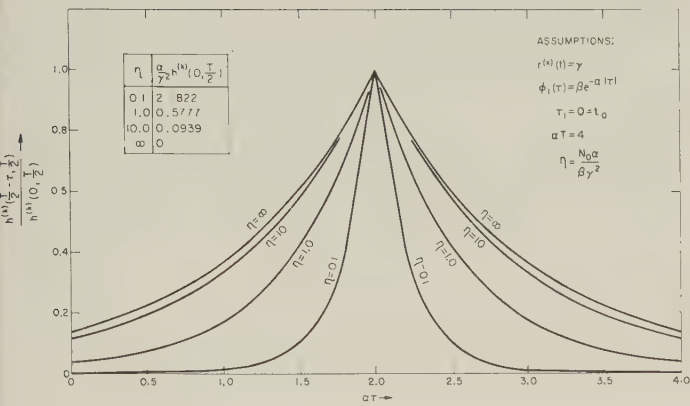


Fig. 3—The optimum filter function  $h^{(k)}(t - \tau, t)$  as a function of the signal-to-noise ratio.

$$\frac{\mu N_0}{2\alpha\beta} h^{(k)}(t - \tau, t)$$

$$\left[ = \frac{I_\nu(\hat{s}_t)I_\nu(\hat{s}_{t'})K_{\nu-1}(\hat{s}_0)K_{\nu+1}(\hat{s}_t) + K_\nu(\hat{s}_t)K_\nu(\hat{s}_{t'})I_{\nu-1}(\hat{s}_0)I_{\nu+1}(\hat{s}_t) + I_{\nu+1}(\hat{s}_t)K_{\nu-1}(\hat{s}_0)[I_\nu(\hat{s}_{t'})K_\nu(\hat{s}_{t'}) + K_\nu(\hat{s}_{t'})I_\nu(\hat{s}_{t'})]}{I_{\nu-1}(\hat{s}_0)K_{\nu+1}(\hat{s}_t) - I_{\nu+1}(\hat{s}_t)K_{\nu-1}(\hat{s}_0)} \right] + \begin{cases} I_\nu(\hat{s}_{t'})K_\nu(\hat{s}_{t'}); & \tau' \leq t' \\ K_\nu(\hat{s}_{t'})I_\nu(\hat{s}_{t'}); & \tau' \geq t' \end{cases} \quad (72)$$

TABLE I  
THE BIAS  $\epsilon(\bar{N}_0)$

$\bar{N}_0$	$\alpha T = 0$	$\alpha T = 0.4$	$\alpha T = 1.0$	$\alpha T = 2.0$	$\alpha T = 4.0$	$\alpha T = 40$	$\alpha T = 400$
0.01	0.046151205	0.103626	0.145791	0.192107			0.900
0.02	0.078636513	0.148721	0.203217	0.262492	0.339938	0.699	0.9444
0.05	0.15222612	0.236536	0.308192	0.384795	0.480307	0.82993	0.97620
0.09	0.2244710	0.313893	0.394407	0.479353	0.580928		
0.10	0.23978953	0.329641	0.411328	0.497296	0.59922	0.901	0.987817
0.11	0.25427974	0.344375	0.426980	0.513714	0.615727		
0.20	0.35835189	0.446570	0.531190	0.618650	0.715968	0.9449	0.9938344
0.50	0.54930615	0.622932	0.696052	0.769268	0.843362	0.97646	
1.0	0.69314718	0.748831	0.804304	0.858203	0.90923		
2.0	0.81093022	0.848083	0.884701	0.919345	0.95048		
5.0	0.9116078	0.930115	0.948112	0.964696			
10.0	0.9531018	0.963169	0.972900	0.981770	0.98940		
20.0	0.9758032	0.981062	0.986143	0.990712	0.99484		

$$\left. \begin{aligned} \tan gR_l &= \frac{2R_l}{R_l^2 - 1} \\ g &= \alpha T \end{aligned} \right\} \quad (67)$$

Using (66) in (25), with each  $\lambda_i^{(k)}$  counted twice, as explained in connection with (56), values of the bias

$$\epsilon(\bar{N}_0) = \bar{N}_0 \log_e F_k^{1/2} \quad (68)$$

have been computed for various  $\bar{N}_0$  and  $g$ , where

$$\left. \begin{aligned} \bar{N}_0 &= N_0/\bar{E}_s \\ \bar{E}_s &= \phi_1(0) \int_{t_0}^{t_1} [r^{(2)}(t - \tau_1)]^2 dt \end{aligned} \right\} \quad (69)$$

$\bar{E}_s$  is the average signal energy received during the observation interval when  $s_k$  is encoded. For the special case under consideration,

$$\bar{N}_0 = \frac{N_0}{\beta\gamma^2 T} \quad (70)$$

The values of  $\epsilon(\bar{N}_0)$  are presented in Table I and Fig. 4 (below) but details of the computation, such as securing of good convergence for (25), will be found elsewhere.<sup>10</sup>

Solution for Exponential  $r^{(k)}(t)$  and Exponential  $\phi_1(\tau)$ :

For

$$r^{(k)}(t - \tau_1) = \gamma e^{-\mu t}; \quad \mu \text{ positive} \quad (71)$$

and  $\phi_1(\tau)$  as in (62), (54), and (56) can be solved by the methods of Youla<sup>19</sup> and Juncosa.<sup>28</sup> The integral equations are converted into second-order differential equations using Youla's method, and Juncosa's transformation is then applied to yield Bessel's form. The integral equations are thus solved by fitting Bessel functions to appropriate boundary conditions. We find

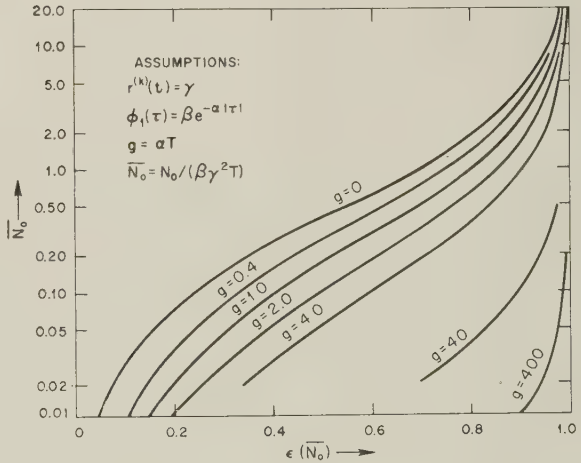


Fig. 4—The bias constant  $\epsilon(\bar{N}_0)$  as a function of the signal-to-noise ratio.

<sup>28</sup> M. L. Juncosa, "An integral equation related to Bessel functions," *Duke Math. J.*, vol. 12, pp. 465-471; 1945.

where

$$\nu = \alpha/\mu \quad (73)$$

$$\hat{s}_\sigma = \nu e^{-\mu\sigma} \sqrt{\frac{2}{\eta}} \quad (74)$$

$$t' = t - t_0; \quad \tau' = \tau - t_0 \quad (75)$$

and  $I_\nu(x)$  and  $K_\nu(x)$  are modified Bessel functions of the first and second kinds, respectively, of order  $\nu$ . As  $\mu \rightarrow 0$ , (72) must approach (63) as a limit. In Fig. 5 (next page), (63) and (72) are compared under similar conditions to illustrate the effect of the form of  $r^{(k)}(t)$  on the shape of  $h^{(k)}(t - \tau, t)$ . As  $\eta \rightarrow \infty$ , the shape of  $h^{(k)}(t - \tau, t)$  converges to that of  $\phi_1(t - \tau)$  irrespective of the form of  $r^{(k)}(t)$ , in agreement with the result (49) for large  $N_0$ .

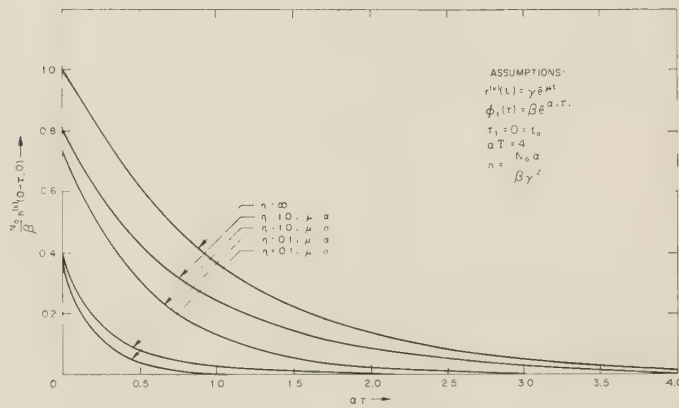


Fig. 5—Effect on the optimum filter function  $h^{(k)}(t - \tau, t)$  of varying  $r^{(k)}(t)$ .

In connection with the use of  $h^{(k)}(t - \tau, t)$  as a least-mean-square-error estimator of  $y_{s1}(t)$  or  $y_{c1}(t)$ , as discussed in "Optimum Detection for a Single Process Random Signal (or a Single-Scatter Path)," it is of interest to examine this minimum error  $E(t)$ . Using (54), (57), and (60), we find

$$E(t) = \overline{[y_{s1}(t) - I_1(t)]^2} = \overline{[y_{c1}(t) - I_2(t)]^2} = N_0 h^{(k)}(0, t), \quad (76)$$

a result valid for all  $r^{(k)}(t)$  and  $\phi_1(\tau)$ . In Fig. 6, the  $E(t)$  are plotted for the two cases considered in this section, under similar conditions. For the case of constant  $r^{(k)}(t)$ , there is a rise in error near the end points of the observation interval, produced by the reduction of the observable neighborhood in which  $y_1(t)$  is approximately constant. For exponentially decaying  $r^{(k)}(t)$ , this effect is present at the start of the observation interval, but as  $r^{(k)}(t)$  decays  $E(t)$  naturally increases.

The solution of (56) for exponential  $r^{(k)}(t)$  and  $\phi_1(\tau)$  may be found by the method used in arriving at (72). The eigenvalues are

$$\lambda_l^{(k)} = \frac{2\alpha\beta\gamma^2}{\mu^2 R_l^2} \quad (77)$$

where the  $R_l$  are the positive roots of

$$J_{\nu+1}(e^{-\mu T} R_l) N_{\nu-1}(R_l) - N_{\nu+1}(e^{-\mu T} R_l) J_{\nu-1}(R_l) = 0. \quad (78)$$

$J_\nu(x)$  and  $N_\nu(x)$  are Bessel functions of the first and second kinds, respectively, of order  $\nu$ . For  $\nu = 1/2$  and  $3/2$ , (78) becomes a trigonometric equation like (67). When  $T = \infty$ , (78) reduces to Kac and Siebert's solution,

$$J_{\nu-1}(R_l) = 0, \quad (79)$$

which assumes simple trigonometric forms for  $\nu = 1/2$ ,  $3/2$ ,  $5/2$ , and  $7/2$ . For  $\nu = 1/2$ , and  $3/2$ , the  $R_l$  are simply

$$R_l = \begin{cases} \frac{\pi}{2}(2l-1); & \nu = 1/2 \\ \pi l; & \nu = 3/2 \end{cases}; \quad T = \infty \quad (80)$$

and the  $\epsilon(\bar{N}_0)$  of (68) may, from (25), be found in closed form:<sup>29</sup>

$$\epsilon(\bar{N}_0) = \begin{cases} \bar{N}_0 \log_e \prod_{l=1}^{\infty} \left[ 1 + \frac{8}{\bar{N}_0 \pi^2 (2l-1)^2} \right] \\ = \bar{N}_0 \log_e \cosh \sqrt{2/\bar{N}_0}; \nu = 1/2 \\ \bar{N}_0 \log_e \prod_{l=1}^{\infty} \left[ 1 + \frac{6}{\bar{N}_0 \pi^2 l^2} \right] \\ = \bar{N}_0 \log_e \frac{\sinh \sqrt{6/\bar{N}_0}}{\sqrt{6/\bar{N}_0}}; \nu = 3/2 \end{cases} T = \infty \quad (81)$$

where, from (69),

$$\bar{N}_0 = \frac{2\mu N_0}{\beta \gamma^2}. \quad (82)$$

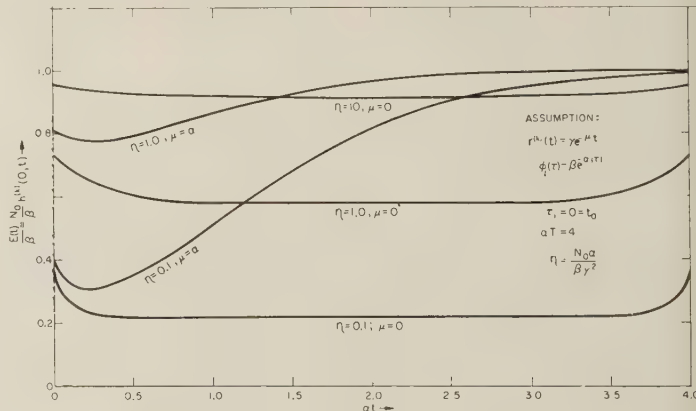


Fig. 6—Mean-square-estimation error  $E(t)$  as a function of  $r^{(k)}(t)$ .

## CONCLUSION

Although complete solution of the integral equations (22) has been possible only in a few special cases, it appears that the optimum filter characteristics do not

<sup>29</sup> P. Franklin, "Treatise on Advanced Calculus," John Wiley & Sons, Inc., New York, N. Y., pp. 463, 473 (48); 1940.

depend critically on the parameters of the systems under consideration. For example, Figs. 5 and 3, respectively, indicate that the shape of  $h^{(k)}(t - \tau, t)$  does not change radically with variations in either  $r^{(k)}(t)$  or the strength of the noise. The  $\phi_p(\tau)$  could probably assume a wide variety of forms other than simple exponential, without markedly altering the general behavior of  $h^{(k)}(t - \tau, t)$ . Similar observations would apply to bias constants  $\epsilon(N_0)$ .

It is also likely that optimum receiver realizations arrived at by using assumption of Gaussian statistics will

perform well for a much broader class of random processes.

In the case of the single-scatter-path channel, a functional form for the optimum detector has been found exactly, although the exact filter functions and biases to employ in the various elements have been established only in two special cases. The general result that the optimum receiver consists of an estimator-correlator combination is of particular significance, and even appears to apply to some extent for multipath channels when the noise is appreciable, as shown by Fig. 2.

# A Coincidence Procedure for Signal Detection\*

MISCHA SCHWARTZ†

**Summary**—A coincidence method of detecting signal in the presence of noise is compared to the statistically optimum Neyman-Pearson procedure utilizing signal integration and threshold detection. In this coincidence procedure a specified number of the fixed group of successive pulses are required to exceed a voltage threshold level. The analysis is carried out for the case of constant-amplitude signals only and the results indicate that the best possible coincidence method requires about 1.4 db more power than the Neyman-Pearson method.

## INTRODUCTION

THE NEYMAN-PEARSON statistical theory of testing hypotheses has recently been applied to the problem of the detection of pulsed signals in the presence of noise.<sup>1-3</sup>

The detection procedure developed on the basis of the Neyman-Pearson theory consists of requiring the sum of  $K$  detected voltage pulses to exceed a specified threshold level. This level is determined by the allowable probability,  $P_{nbK}$ , that noise in the absence of signal might exceed the level and be erroneously detected as a signal. The probability,  $P_{sbK}$ , that signal plus noise will exceed the level (after summing or integration) and be identified as signal is then dependent upon signal-to-noise ratio.

The Neyman-Pearson threshold detection scheme is presumably optimum in the sense that it requires minimum signal-to-noise ratio for given  $P_{nbK}$  and  $P_{sbK}$ . The obvious question arises, however, that if this procedure is optimum how much better is it than other possible procedures? Do other procedures exist which do not perhaps do the

required job as efficiently, but which may be warranted anyway because of simplicity and economy of equipment required?

A general answer to these basic questions will not be attempted here. Instead an alternative statistical detection procedure will be analyzed and compared to the Neyman-Pearson optimum approach. This can then serve as a first step in the process of finding answers to the above questions.

The procedure to be discussed is very simple in concept and consists basically of a coincidence threshold scheme for the detection of signals in the presence of noise. The threshold level of the previous optimum integration method is retained, but now a specified number of the individual (nonintegrated)  $K$  pulses are required to exceed the level to be called a signal. It is felt that equipment-wise, this coincidence procedure may be at least as amenable to development as the optimum integration scheme, a counter (or counters) with set and reset gates being a possibility for a practical detection system of this nature. The counter would be adjusted to indicate signal detection if  $n$  pulses of the group of  $K$  ( $n \leq K$ ) succeeded in passing the threshold level and registering in the counter. In a radar system a different counter would be required for each range element within the range gate of the system, or else provision would have to be made to store the information contained in the different range elements, implying a storage tube as in the integration case. At the end of a time equal to  $K$  repetition periods the counters would be reset to zero and the procedure again repeated. This scheme compares with the storage tube and associated circuitry, or other devices required for signal integration.

Since such a coincidence detection method might be fairly readily applied practically, it is of interest to determine analytically how efficiently this procedure functions for the purpose of signal detection. The following analysis is aimed at answering this question.

\* Manuscript received by the PGIT, March 28, 1956. Abstracted from Chap. 4 of an unpublished Ph.D. dissertation, Harvard Univ., Cambridge, Mass., May 1, 1951. The work was done under a scholarship grant of the Sperry Gyroscope Co., Great Neck, N. Y.

† Brooklyn Polytech. Inst., Brooklyn, N.Y.

<sup>1</sup> M. Schwartz, "A Statistical Approach to the Automatic Search Problem," Ph.D. dissertation, Harvard Univ., 1951.

<sup>2</sup> H. V. Hance, "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Sc.D. dissertation, Mass. Inst. Tech., Cambridge, Mass., 1951.

<sup>3</sup> D. L. Drukey, "Optimum Techniques for Detecting Pulse Signals in Noise," presented at the IRE Natl. Conv., New York, N. Y., March 4, 1952.

To facilitate matters it is assumed that signals when received are all of the same constant amplitude.<sup>1</sup> The case of fluctuating or unequal signals will not be considered in the coincidence analysis in this paper.

### COINCIDENCE ANALYSIS

With the assumption of constant-amplitude signals each pulse among  $K$  successive voltage pulses will have the same probability,  $p$ , of exceeding a specified threshold level. When the pulse is that due to noise,  $p$  is exactly  $P_{nb}$  (the probability of one noise pulse exceeding the level) of the Neyman-Pearson threshold scheme. Similarly, for signal plus noise,  $p$  is identically  $P_{sb}$ . The probability  $P$  that exactly  $n$  out of  $K$  pulses will exceed the level is then given by the binomial distribution as

$$P = {}_K C_n p^n (1 - p)^{K-n} \quad (1)$$

where

$${}_K C_n = \frac{K!}{n!(K - n)!}$$

represents the number of ways in which the required  $n$  successes out of  $K$  trials can occur, and  $(1 - p)$  is the probability of failure (*i.e.*, the voltage fails to exceed the threshold level).

The presence of a signal will, however, also be indicated in cases where more than  $n$  voltages of the  $K$  received would have exceeded the level had the counter been allowed to count beyond the first  $n$  pulses. The probability,  $P_s$ , that a signal will be detected by this coincidence procedure is thus given by the cumulative binomial distribution,

$$P_s = \sum_{m=n}^K {}_K C_m P_{sb}^m (1 - P_{sb})^{K-m}, \quad (2)$$

the sum of the mutually exclusive probabilities that  $n$  or  $(n + 1)$  or \*\*\* $K$  signal-plus-noise pulses will each exceed the level.

Similarly,  $P_n$ , the probability that noise will be erroneously called a signal is given by

$$P_n = \sum_{m=n}^K {}_K C_m P_{nb}^m (1 - P_{nb})^{K-m}. \quad (3)$$

( $P_{sb}$  and  $P_{nb}$  are the single-voltage probabilities defined previously.)

Eqs. (2) and (3) are the basic equations of the coincidence procedure. With  $n$  and  $K$  fixed, and  $P_n$  specified,  $P_{nb}$  is determined from (3).

In this case  $P_{nb}$  is simply the probability that a single noise sample, after envelope detection, will exceed a fixed threshold level. With Gaussian noise assumed at the input to a square-law detector,  $P_{nb}$  is given by<sup>1</sup>

$$P_{nb} = e^{-q} \quad (4)$$

with

$$q \equiv \frac{b}{2a\psi_0}$$

$b$ —the threshold level in volts,  
 $a$ —a detector constant, and  
 $\psi_0$ —the mean-squared noise voltage.

Setting  $P_s$  at a desired value (as operationally determined, *e.g.*, 90 or 99 per cent,) the required value for  $P_{sb}$  is then found from (2).  $P_{sb}$  is here the probability that a single signal-plus-noise sample will exceed the threshold level.

The equation for  $P_{sb}$  was derived by Rice in his classic paper.<sup>4</sup> Combining (4) and Rice's results,  $P_{sb}$ ,  $P_{nb}$ , and mean power signal-to-noise ratio,  $s^2$ , can be shown to be related by:

$$P_{sb} = \int_{\sqrt{-\log e P_{nb}}}^{\infty} 2ye^{-(y^2 + s^2)} I_0(2sy) dy. \quad (5)$$

$I_0(x)$  is the modified Bessel function of first kind and zero order. (Note that this equation is independent of the type of envelope detector used.)

Eq. (5) has been plotted in Fig. 1, using the curves available in Rice's article.

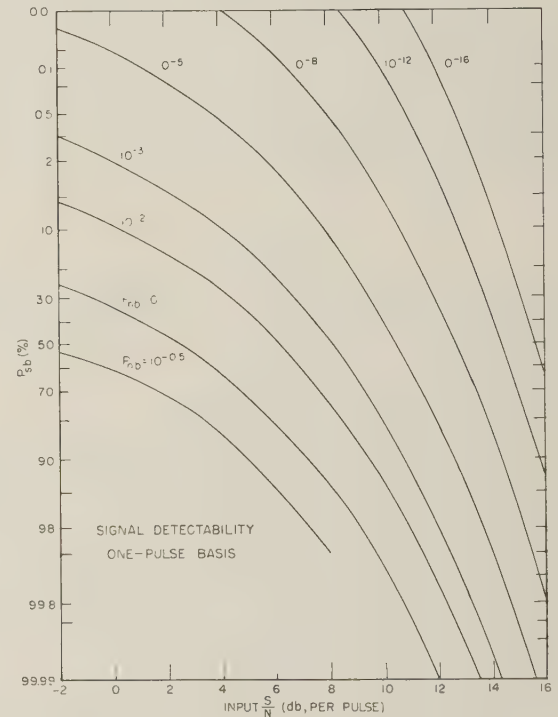


Fig. 1

Since  $P_s$  and  $P_n$  here are analogous to  $P_{sbK}$  and  $P_{nbK}$ , respectively, in the signal integration procedures of detection, the coincidence and integration procedures may be compared, for a given  $K$ , by determining the input signal-to-noise ratio required for each scheme (with  $P_s = P_{sbK}$ ,  $P_n = P_{nbK}$ ).

### OPTIMUM NUMBER OF COINCIDENCE COUNTS

Thus far there has been no discussion as to the particular choice for the parameter  $n$ . For a given value of  $K$ , and

<sup>4</sup> S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 102-103; January, 1945.

with  $P_n$  and  $P_s$  specified, that  $n$  (or range of values of  $n$ ) should be chosen which results in a minimum signal-to-noise ratio required per pulse. That such an optimum value for  $n$  does exist can be shown qualitatively.

Thus, for  $n$  small the very high threshold levels needed to prevent noise from being detected as signal force large signal powers to be needed, while for  $n$  approaching  $K$  the requirement that  $n$  signal plus noise voltages exceed the level is stringent enough to require large signal powers again.

It has not been found possible to find the optimum  $n$  analytically. Instead a point-by-point graphical procedure has been followed. For this purpose the National Bureau of Standards' tables of the cumulative binomial distribution have been used,<sup>5</sup> these tables giving  $P_{sb}$  directly when  $P_s$ ,  $K$  and  $n$  in (2) are specified. These tables cover part of the range of values of  $P_n$ , ( $10^{-5} \geq P_n \geq 10^{-7}$ ), the remaining range required,  $10^{-7}$  to  $10^{-13}$ , having been calculated from (3). Using these tables first to obtain  $P_{nb}$  and  $P_{sb}$ , and then Fig. 1 to find the required input signal-to-noise ratio per pulse, it was possible to plot this signal-to-noise ratio as a function of  $n$  ( $1 \leq n \leq K$ ) for various values of  $K$ ,  $P_n$ , and  $P_s$ .

The results are shown in Figs. 2 to 6 on the next page, for  $K = 3, 5, 10, 30$ , and  $49$ , respectively, and for the various indicated combinations of  $P_n$  and  $P_s$ . The points indicated on the curves are at the values of  $n$  for which the calculations were made.

Since interpolation between the curves of Fig. 1 was necessary, signal-to-noise ratios could not be found too accurately, so that the results in Figs. 2 to 6 are accurate only to within a few tenths of a  $db$ , good enough, however, for the purposes of this analysis. This accounts for the apparently peculiar-looking shapes of some of the curves in these figures.

The regions of optimum  $n$  are clearly shown, however, and are seen to be fairly broad for  $K$  large enough.

With  $K$  fixed and greater than 10, the region of optimum  $n$  seems to shift slowly toward smaller values of  $n$  as  $P_s$  decreases or  $P_n$  increases. Because of the scattering of the points plotted, however, it is not possible to positively verify this conclusion. If such a shift does occur it is small enough so that for  $10^{-10} \leq P_n \leq 10^{-5}$  and 50 per cent  $\leq P_s \leq 90$  per cent the broad region of optimum  $n$  can be considered to be roughly independent of  $P_n$  and  $P_s$ , (for a particular  $K$ ).

The trend of decreasing optimum  $n$  with increasing  $K$  is definitely noticeable on the other hand. The exact dependence of optimum  $n$  on  $K$  has not been determined analytically, but judging from the graphs the optimum value of  $n$  seems to be roughly proportional to  $\sqrt{K}$ .

From the results for the values of  $K$  plotted, the region of optimum  $n$  (minimum input signal-to-noise ratio within 0.2  $db$ ) may be written empirically with good enough accuracy as

$$\text{opt. } n \doteq 1.5\sqrt{K}. \quad (6)$$

This is valid for  $10^{-10} < P_n < 10^{-5}$ , 50 per cent  $< P_s < 90$  per cent.

## RESULTS OF ANALYSIS

Using the optimum values of  $n$  the results of the coincidence detection procedure have been plotted in Figs. 7 and 8 on the next page, as the curves labeled "optimum coincidence." (These curves thus give the minimum signal-to-noise ratio required for different values of  $K$ ,  $P_n$  and  $P_s$ . The minimum signal-to-noise ratios have been obtained from Figs. 2 to 6.)

These results are there compared with results obtained for the Neyman-Pearson threshold integration procedure<sup>1</sup> ("video integration"), in which  $P_{sbK}$  and  $P_{nbK}$  have been set equal to  $P_s$  and  $P_n$  respectively.

Also plotted are curves for the particular case of total or complete coincidence, all  $K$  pulses being required to exceed the threshold level for detection; "linear integration" curves (an ideal situation corresponding to coherent detection); and curves for a system integrating only as efficiently as  $\sqrt{K}$ . (Here the required signal-to-noise ratio decreases as the square-root of the samples added. This has been found experimentally to be the performance capability of a human operator detecting signals on a screen.)

Figs. 7 and 8 indicate that the "optimum coincidence" curves are very nearly parallel to the Neyman-Pearson video integration curves (at least up to  $K = 49$ , the largest value of  $K$  taken), requiring 1.3 to 1.5  $db$  more power for the same system performance with  $K \geq 10$ , the additional power needed decreasing with smaller values of  $K$ . (All threshold detection curves of course coincide at  $K = 1$ ). This coincidence "loss" of about 1.4  $db$  over video integration is very nearly the same for the two cases plotted,  $P_s = 90$  per cent and 50 per cent ( $P_n = 10^{-10}$ ), and can be shown to be about the same value for  $P_n = 10^{-5}$ .

These results mean that the coincidence procedure might in some cases be preferred over the integration procedure if the circuitry needed were less complex.

It is noteworthy to remark that although the optimum coincidence procedure is 1.4  $db$  worse than the integration method, it is still much better in performance than a system integrating only as efficiently as  $\sqrt{K}$ .

For  $K$  small enough (less than about 30 for the values of  $P_n$  and  $P_s$  in Figs. 6 and 7) the complete coincidence procedure is also more efficient than the  $\sqrt{K}$  integration.

It is well to remember, however, that the above conclusions as to the comparative efficiency of the two detection schemes considered are strictly correct only under the given assumption of equal-amplitude non-fluctuating signal echoes.

Although the effect of signal fluctuation on the Neyman-Pearson detection procedure has been analyzed,<sup>1</sup> similar calculations have not been carried out for the coincidence method.

<sup>5</sup> National Bureau of Standards, "Tables of the binomial probability distribution," *Appl. Math. Series*, no. 6; January, 1950.

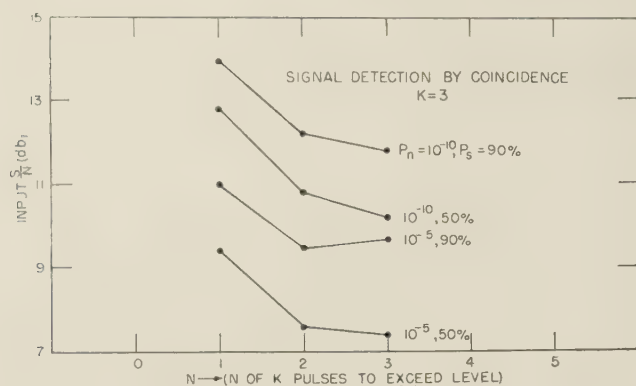


Fig. 2

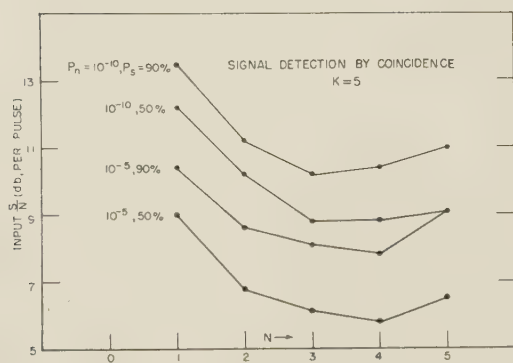


Fig. 3

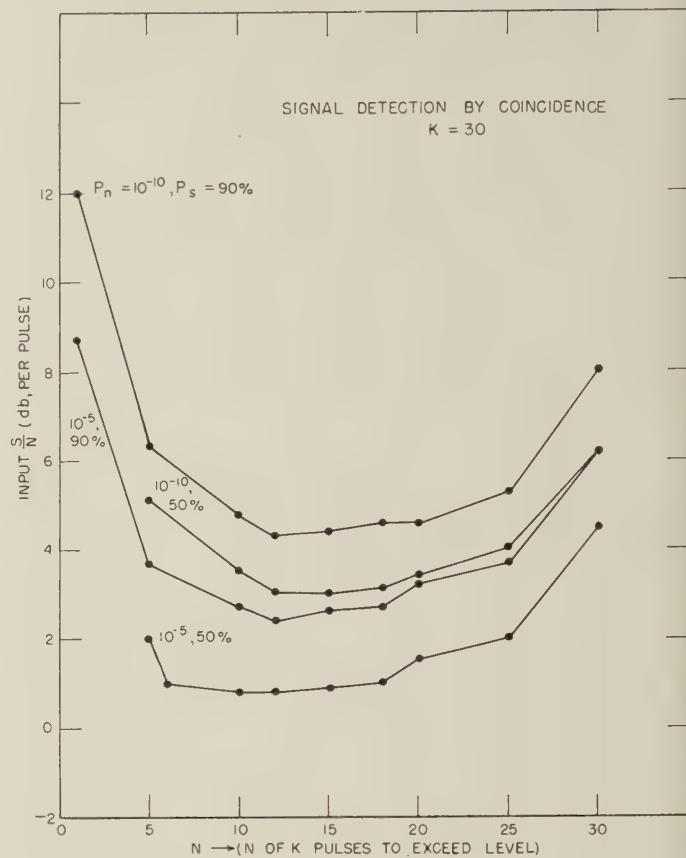


Fig. 5

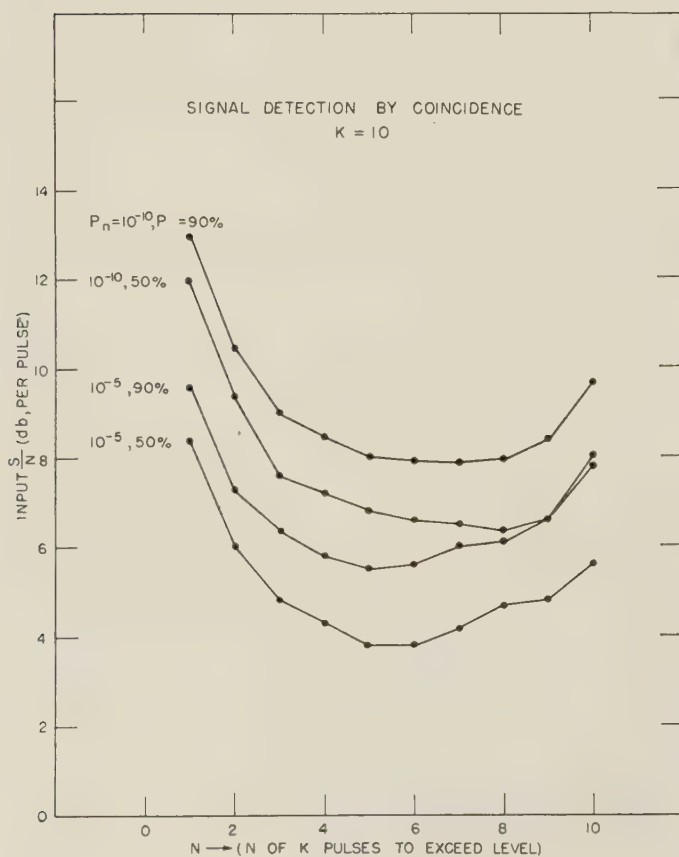


Fig. 4

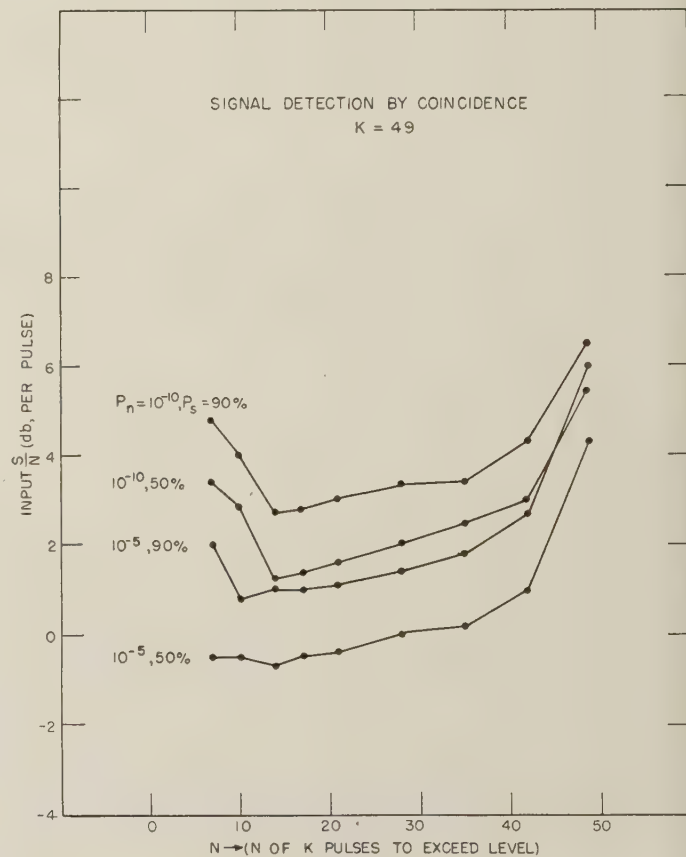


Fig. 6

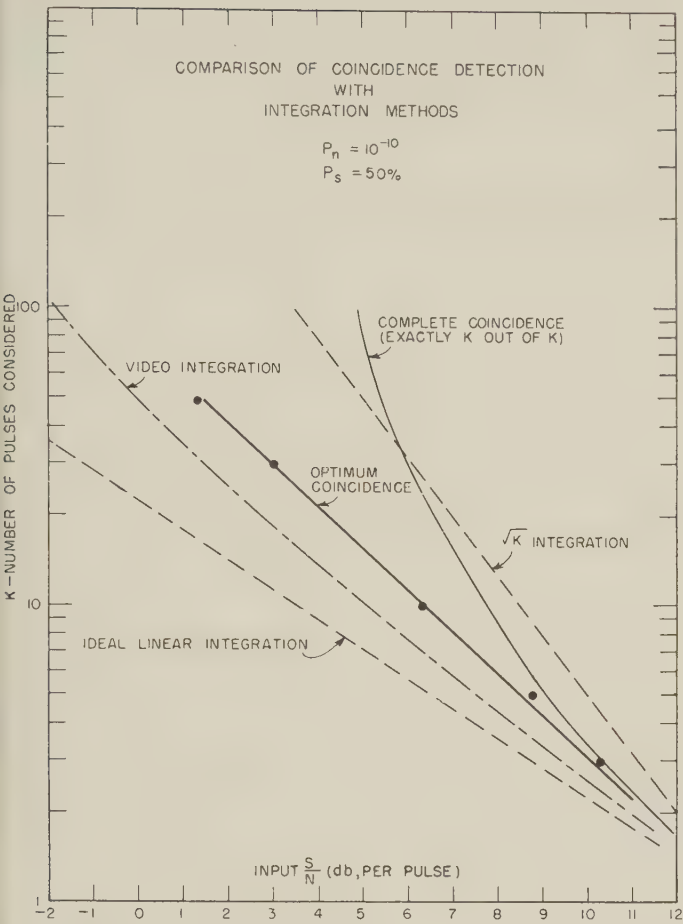


Fig. 7

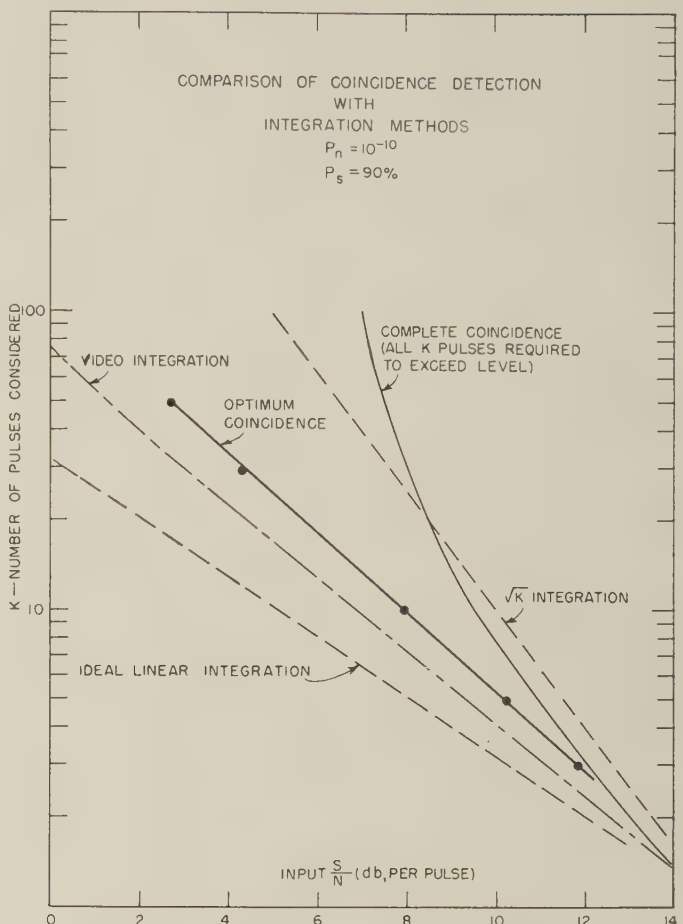


Fig. 8

# Some General Aspects of the Sampling Theorem\*

D. L. JAGERMAN† AND L. J. FOGEL‡

**Summary**—The sampling theorem is recognized as an interpolation formula. Starting from the Lagrange Polynomial, this theorem is developed under conditions which are of broader applicability than those usually stated. Such a view point indicates the essential unity of temporal and frequency domain application. It will also be shown that the theorem is applicable as an exact interpolation formula throughout the complex plane. The basic theorem is extended to include sampling of the first derivative of the function. The concept of band-limited functions is introduced through use of Fourier-Stieltjes representations. This is then shown to be subsumed under the general class of functions which is considered in connection with the interpolation theorems developed. This approach, as presented, readily leads to the establishment of many sampling theorems. It is hoped that this paper will aid the formulation of particularly applicable sampling theorems for specific problems.

## INTRODUCTION

IN SOME communication systems the discrete time samples of data which are received concerning the original signal must be smoothed so as to yield a continuous replica of the original signal. The choice of the smoothing process is normally dependent upon the amount of reliance which is placed on the sample information. For example, it may prove of value to average each received pulse of the signal so as to obtain a single sample value, then accomplish smoothing by application of the sampling theorem, that is, the use of the convergent cardinal series which yields all intermediate values of the signal. This paper will indicate the relation this smoothing process has with respect to interpolation theory and in this manner emphasize the required conditions which must be fulfilled for valid application of the theorem.

\* Manuscript received by the PGIT, May 31, 1956.  
† Stavid Engineering, Inc., Plainfield, N. J.  
‡ Convair, San Diego, Calif.

Practical application of the extended sampling theorem in the complex domain may serve to facilitate complex spectral analysis based upon the knowledge of system operation at some discrete points. It appears necessary to recognize some of the fundamental relationships which underlie the derivation of this theorem so as to permit proper choice of interpolation technique for each specific problem.

### DISCUSSION

Let the available discrete data be represented by  $f(t_0)$ ,  $f(t_1)$ ,  $f(t_2)$ ,  $f(t_3)$ ,  $\dots$   $f(t_n)$ , and for simplicity of notation,

$$f(t_n) = f_n. \quad (1)$$

It is often desired to construct a continuous interpolate from such discrete data which may be obtained as samples of an original continuous signal. If reliance is placed on the correctness of the discrete information, then it becomes reasonable to require the interpolate to pass through the sample values. (Alternate smoothing processes could also be chosen so as to minimize the mean square error, to minimize the maximum error, or to improve some other imposed metric or system characteristic. The following discussion will be restricted to errorless interpolation at the sample points achieved by selection of a proper polynomial.)

A polynomial of the  $n$ th degree is exactly specified by  $n + 1$  points and has  $n$  zeros (including multiplicity of zeros). Lagrange utilized a general interpolation polynomial<sup>1</sup> of the form

$$P_n(t) = f_0 L_0^n(t) + f_1 L_1^n(t) + \dots + f_n L_n^n(t) \quad (2)$$

where  $L_i^n(t)$  is the Lagrangean Coefficient defined by

$$L_i^n(t) = \frac{(t-t_0)(t-t_1) \dots (t-t_{i-1})(t-t_{i+1}) \dots (t-t_n)}{(t_i-t_0)(t_i-t_1) \dots (t_i-t_{i-1})(t_i-t_{i+1}) \dots (t_i-t_n)}. \quad (3)$$

Note that this cofactor has the following "selective" property;

$$\begin{aligned} L_i^n(t_j) &= 0, & i &\neq j \\ &= 1, & i &= j \end{aligned} \quad (4)$$

thus each sample point in time will cause all but one of the polynomial terms in (2) to vanish; the remaining term will have a coefficient of unity. In this manner the entire polynomial will agree with the sample data at each of the sample points. Such an interpolation polynomial is clearly unique so that any other interpolation polynomial which yields  $f_i$  at  $t = t_i$  for  $0 \leq j \leq n$  can be exhibited in the form of (2). It may be shown<sup>2</sup> that a scale change applied to the time axis will not affect the Lagrangean Coefficients when the sample points are

equidistant. Fig. 1 illustrates the nature of a typical Lagrangean Coefficient as a function of the continuous argument.

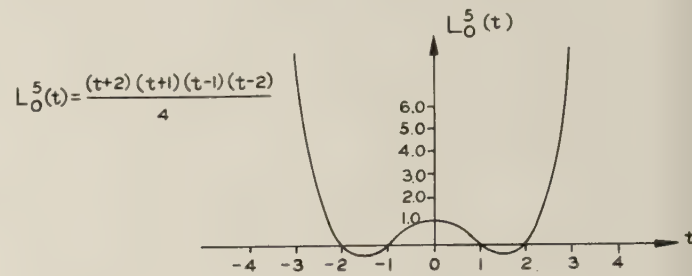


Fig. 1—The Lagrangean coefficient  $L_0^5$ .

It is of interest to compute the error of interpolation. This may be found from the relation

$$f(t) - P_n(t) = \frac{f^{(n+1)}(\tau)}{(n+1)!} (t-t_0)(t-t_1) \dots (t-t_n) \quad (5)$$

where  $\tau$  is a particular time which lies somewhere between the least and the greatest  $t_i$ .<sup>3</sup> By the above remark concerning the uniqueness of polynomials this equation also represents the error obtained in interpolation through a large number of standard interpolation techniques, including the Newton-Gauss and the Newton-Gregory procedures.

Let  $g_n(t)$  be defined by

$$g_n(t) = (t-t_0)(t-t_1)(t-t_2) \dots (t-t_n), \quad (6)$$

then the Lagrangean Coefficient may be expressed in the following compact form

$$L_i^n(t) = \frac{g_n(t)}{(t-t_i)g_n'(t_i)} \quad (7)$$

and the Lagrangean interpolation polynomial (2) becomes

$$P_n(t) = g_n(t) \sum_{i=0}^n \frac{f_i}{(t-t_i)g_n'(t_i)}, \quad (8)$$

which is in the form of a partial fraction expansion. Nothing has been said to this point which necessitates consideration of only a real variable argument; therefore, let  $t$  be replaced by the complex variable  $z$ . Dividing by  $g_n(z)$  yields

$$\frac{P_n(z)}{g_n(z)} = \frac{f_0}{(z-z_0)g_n'(z_0)} + \dots + \frac{f_n}{(z-z_n)g_n'(z_n)}. \quad (9)$$

Examination of the right hand expression shows that it is meromorphic with a simple pole at each sample point. The polynomial  $P_n(z)$  is an entire function; however, the fraction  $P_n(z)/g_n(z)$  is meromorphic to agree with the right-hand side of (9) as is seen from the nature of  $g_n(z)$ .

<sup>3</sup> The time  $\tau$  may be eliminated from explicit consideration by obtaining an estimate  $M$  satisfying  $|f^{(n+1)}(\tau)| \leq M$ , least  $t_i < \tau < \text{greatest } t_i$  then (5) may be written

$$|f(t) - P_n(t)| \leq \frac{M}{(n+1)!} |(t-t_0)(t-t_1) \dots (t-t_n)|.$$

<sup>1</sup> This was first due to Euler who, in a tract entitled "De Eximio Usu Methodi Interpolationum in Serierum Doctrina," had long before obtained a closely analogous expression.

<sup>2</sup> See the appendix.

The function,  $g_n(z)$ , is entire, but vanishes at the sample points. The fraction may be represented by

$$\Phi(z) = \frac{P_n(z)}{g_n(z)}. \quad (10)$$

This leads to a generalization where the number of sample points becomes infinite. A suitable choice of entire function which vanishes at a denumerable infinity of equidistant points is

$$g(z) = \sin \frac{\pi}{h} z, \quad (11)$$

where  $g(z)$  replaces  $g_n(z)$ . The zeros of this function occur at

$$z = jh, \quad j = \dots -3, -2, -1, 0, +1, +2, \dots \quad (12)$$

and so, form a set of equidistant points which lie on a straight line with the central sample point placed at the origin. The right-hand side of (8) rewritten in terms of  $g(z)$  is

$$\sin \frac{\pi}{h} z \sum_{i=-\infty}^{\infty} (-1)^i \frac{f_i}{z - z_i} \quad (13)$$

which is recognized to be the cardinal series. This straight line may have any slope as determined by the argument of the sampling distance  $h$ ;  $h$  being, in general, complex. This paper will be limited to considering only samples determined by a fixed complex  $h$ ; thus, sample points which lie on curved lines are excluded.

The form of (9) suggests the use of contour integration in that it characteristically resembles the sum of residue terms. Accordingly, under the  $g(z)$  chosen above, consider

$$I_n(z) = \frac{1}{2\pi i} \int_{C_n} \frac{f(\zeta)}{(\zeta - z) \sin(\pi/h)\zeta} d\zeta, \quad (14)$$

where the integration is carried out in the complex  $\zeta$ -plane ( $\zeta = \xi + i\eta$  in which  $\xi$  and  $\eta$  are real) and  $z \neq jh$ . The sequence of paths,  $C_n$ , may consist of any closed contours which, successively enclose additional pairs of singularities corresponding to sample points. Since these sample points are restricted to lie upon a straight line, it may prove expeditious to choose  $C_n$  to be a square as in Fig. 2.

It is convenient to introduce the function  $A(y)$  defined by

$$A(y) = \max_{-\infty < x < \infty} |f(ze^{i\varphi})| \quad (15)$$

for any entire function  $f(z)$  and given constant whenever the maximum value exists. The following theorem will now be established.

# Theorem 1

The entire function  $f(z)$  is given by

$$f(z) = \sum_{i=-\infty}^{\infty} f_i \frac{\sin(\pi/h)(z - jh)}{(\pi/h)(z - jh)},$$

in which the cardinal series is uniformly convergent in any finite closed domain of the  $z$ -plane,  $f_i = f(jh)$ ,

$h = |h| e^{i\varphi}$ , if there is a constant  $K$  so that

$$e^{-(\pi/|h|)|y|} A(y) \leq \frac{K}{|y|}, \quad \text{for } |y| \rightarrow \infty.^4$$

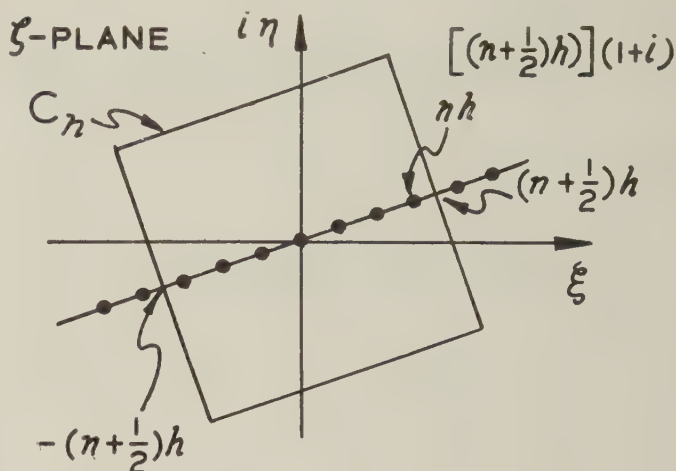


Fig. 2—Contour integration in the  $\zeta$  plane.

To establish the theorem, it is convenient to rotate the path  $C_n$ . Let

$$\tau = \zeta e^{-i\varphi}, \quad (16)$$

then  $I_n$ , (14), may be written as

$$I_n = \frac{e^{i\varphi}}{2\pi i} \int_{D_n} \frac{f(\tau e^{i\varphi})}{(\tau e^{i\varphi} - z) \sin(\pi/|h|)\tau} d\tau. \quad (17)$$

The new path of integration,  $D_n$ , is shown, on the  $\tau$  plane, in Fig. 3. This corresponds to a rotation of the sample points which now occur on the real  $\tau$  axis at the points  $j|h|$ .

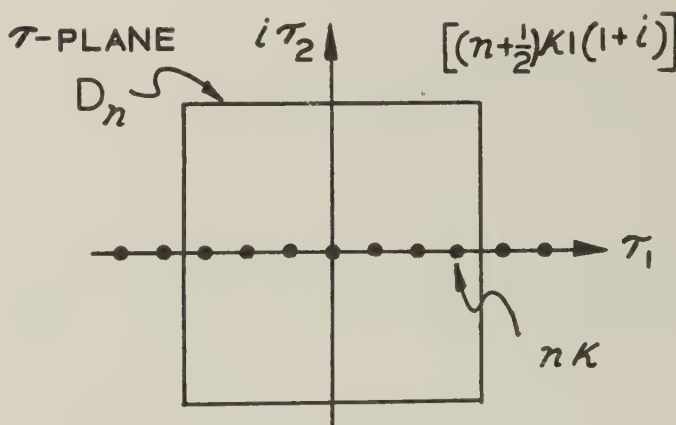


Fig. 3—Contour integration in the  $\tau$  plane.

Denote the integral in (17) taken over the upper horizontal line by  $I_{n,1}$  then

$$|I_{n,1}| \leq \frac{(2n+1)|h|}{2\pi} \max \left| \frac{f(\tau e^{i\varphi})}{(\tau e^{i\varphi} - z) \sin(\pi/|h|)\tau} \right| \quad (18)$$

<sup>4</sup> This notation means  $|y|$  approaches infinity.

in which the maximum is taken on the line. It is assumed also that  $z$  is some point interior to  $D_n$ . Since

$$|\tau e^{i\varphi} - z| \geq (n + \frac{1}{2}) |h| - |z|, \quad (19)$$

it follows that,

$$|I_{n,1}| \leq \frac{(2n+1)|h|}{\pi[(2n+1)|h| - 2|z|]} \max \frac{|f(\tau e^{i\varphi})|}{|\sin(\pi/|h|)\tau|}, \quad (20)$$

also there is, clearly, a constant  $K_1$ , independent of  $n$ , so that

$$\frac{(2n+1)|h|}{\pi[(2n+1)|h| - 2|z|]} \leq K_1. \quad (21)$$

Hence

$$|I_{n,1}| \leq K_1 \max \frac{|f(\tau e^{i\varphi})|}{|\sin(\pi/|h|)\tau|}. \quad (22)$$

An upper bound on

$$\frac{1}{|\sin(\pi/|h|)\tau|}$$

may be obtained as follows:

$$\begin{aligned} \frac{1}{|\sin(\pi/|h|)\tau|} &= \left| \frac{2i}{e^{i(\pi/|h|)\tau} - e^{-i(\pi/|h|)\tau}} \right| \\ &\leq \frac{2}{e^{(\pi/|h|)\tau_2} - e^{-(\pi/|h|)\tau_2}} \leq 4e^{-(\pi/|h|)\tau_2} \end{aligned} \quad (23)$$

in which  $D_n$  has been chosen so that

$$1 - e^{-(2\pi/|h|)\tau_2} \geq \frac{1}{2}. \quad (24)$$

Thus, using (23) and the premise of the theorem,

$$\begin{aligned} |I_{n,1}| &\leq 4K_1 e^{-(\pi/|h|)\tau_2} A(\tau_2) \\ &\leq \frac{4K_1 K}{|\tau_2|}. \end{aligned} \quad (25)$$

Denote the integral along the lower horizontal line by  $I_{n,3}$ , then, in the same manner as above,

$$|I_{n,3}| \leq 4K_2 e^{(\pi/|h|)\tau_2} A(\tau_2) \leq \frac{4K_1 K}{|\tau_2|}. \quad (26)$$

The bounds given in (25) and (26) are uniform in  $\tau_1$ .

Denote the integral taken over the right-hand vertical line by  $I_{n,4}$ , then

$$|I_{n,4}| \leq \frac{1}{2\pi} \int_{-(n+1/2)|h|}^{(n+1/2)|h|} \frac{A(\tau_2)}{|\tau e^{i\varphi} - z| |\sin(\pi/|h|)\tau|} d\tau_2. \quad (27)$$

Since

$$\frac{1}{|\sin(\pi/|h|)\tau|} = \frac{1}{\cosh(\pi/|h|)\tau_2} \leq 2e^{-(\pi/|h|)\tau_2}, \quad (28)$$

it follows that, by the premise of the theorem,

$$\frac{A(\tau_2)}{|\sin(\pi/|h|)\tau|} \leq \frac{2K}{|\tau_2|}. \quad (29)$$

$$|\tau e^{i\varphi} - z| \geq \sqrt{\tau_1^2 + \tau_2^2} - |z|. \quad (30)$$

Thus the integral in (27) converges as  $n \rightarrow \infty$ , and, hence,

$$|I_{n,4}| \leq \frac{1}{\pi} \int_0^\infty \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2. \quad (31)$$

Let

$$\begin{aligned} &\int_0^\infty \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \\ &= \int_0^\eta \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \\ &+ \int_\eta^\infty \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \end{aligned} \quad (32)$$

in which  $\eta > 0$  is arbitrary, then the integral over the range from  $\eta$  to infinity converges uniformly in  $\tau_1$ . It follows that

$$\lim_{|\tau_1| \rightarrow \infty} \int_\eta^\infty \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 = 0. \quad (33)$$

Denote the integral taken on the left-hand vertical line by  $I_{n,2}$ , then, in a similar manner as above, one obtains the same bound for  $|I_{n,2}|$  as that given for  $|I_{n,4}|$  in (31).

Since

$$|I_n| \leq |I_{n,1}| + |I_{n,2}| + |I_{n,3}| + |I_{n,4}| \quad (34)$$

an estimate can be obtained for  $|I_n|$ . Choose  $\eta$  so that

$$\begin{aligned} &\int_0^\eta \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \\ &\leq \int_0^\eta \frac{A(\tau_2)}{(\sqrt{a^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \leq \frac{\epsilon}{6} \end{aligned} \quad (35)$$

in which  $\tau_1 > a > |z|$  and  $\epsilon > 0$  is arbitrarily assigned. Choose  $D_n$  so that

$$|I_{n,1}| \leq \frac{\epsilon}{6}, \quad (36)$$

$$|I_{n,3}| \leq \frac{\epsilon}{6}, \quad (37)$$

$$\int_\eta^\infty \frac{A(\tau_2)}{(\sqrt{\tau_1^2 + \tau_2^2} - |z|) \cosh(\pi/|h|)\tau_2} d\tau_2 \leq \frac{\epsilon}{6}, \quad (38)$$

then

$$|I_n| \leq \epsilon \quad (39)$$

and therefore

$$\lim_{n \rightarrow \infty} I_n = 0. \quad (40)$$

Evaluation of  $I_n$ , (13), by residues will now yield the required result. The singularity at  $\zeta = z$  has the residue

$$\frac{f(z)}{\sin(\pi/h)z}. \quad (41)$$

The singularities introduced by  $\sin(\pi/h)\zeta$  are all simple poles. To evaluate the residue at  $\zeta = jh$ , let

$$W = \zeta - jh \quad (42)$$

then

$$\frac{f(\zeta)}{(\zeta - z)\sin(\pi/h)\zeta} = (-1)^j \frac{f(W + jh)}{(W + jh - z)\sin(\pi/h)W}. \quad (43)$$

The coefficient of  $1/W$  in the Laurent expansion about  $W = 0$  is the required residue. Since

$$f(W + jh) = f_i + Wf'_i + \dots, \quad (44)$$

$$\frac{1}{W + jh - z} = \frac{1}{jh - z} - \frac{W}{(jh - z)^2} + \dots, \quad (45)$$

$$\frac{1}{\sin(\pi/h)W} = \frac{1}{(\pi/h)W} + \frac{1}{6}\left(\frac{\pi}{h}\right)W + \dots, \quad (46)$$

the residue is readily found to be

$$\frac{h}{\pi}(-1)^j \frac{f_i}{jh - z}. \quad (47)$$

Hence

$$\frac{f(z)}{\sin(\pi/h)z} + \frac{h}{\pi} \sum_{i=-\infty}^{\infty} (-1)^j \frac{f_i}{jh - z} = 0 \quad (48)$$

and the theorem follows.

The form of the cardinal series is that of a convolution with kernel  $\frac{\sin(\pi/h)z}{\pi/h)z}$ . Restricting  $z$  to the real axis and  $h$  to be positive, it is of interest to consider the Fourier transform or frequency spectrum of this kernel. One has

$$\begin{aligned} F\left[\frac{\sin(\pi/h)x}{(\pi/h)x}\right] &= \int_{-\infty}^{+\infty} e^{-i\omega x} \frac{\sin(\pi/h)x}{(\pi/h)x} dx \\ &= h, \quad |\omega| < \frac{\pi}{h} \\ &= 0, \quad |\omega| > \frac{\pi}{h}. \end{aligned} \quad (49)$$

Reference is made to Fig. 4. The frequency spectrum is

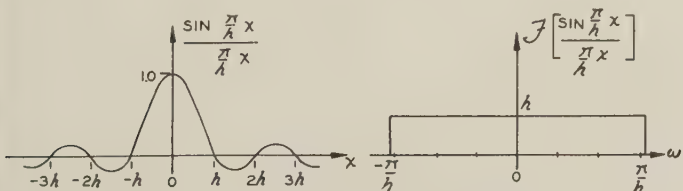


Fig. 4—Cardinal series kernel and its transform.

that of the ideal lowpass filter; however, the kernel is employed in a convolution sum rather than integral. Use of a low-pass filter, therefore, requires modification of interpretation. Let a sequence of equidistant, narrow pulses be used of repetition rate  $1/h$ , width  $\tau$ , and height

$f_i$ , then the output of a low-pass filter with this sequence as input closely approximates<sup>5</sup>

$$\tau \sum_{i=-\infty}^{\infty} f_i \frac{\sin(\pi/h)(z - jh)}{(\pi/h)(z - jh)}. \quad (50)$$

Thus, if  $f(z)$  satisfies the conditions of theorem 1, the output of the filter closely approximates  $\tau f(z)$ .

Interpolation using  $f_i$  and  $f'_i$ , that is

$$\left. \frac{df(z)}{dz} \right|_{z=jh},$$

may be readily accomplished using the guiding principle introduced above. The crux of the matter lies in the proper choice of entire function  $g(z)$ . The zeros of  $g(z)$  must be as before, *i.e.*, at  $z = jh$ , however, in order to introduce  $f'_i$  every zero should be double. Thus a proper choice is

$$g(z) = \sin^2 \frac{\pi}{h} z \quad (51)$$

and the appropriate integral to consider is

$$I_n = \frac{1}{2\pi i} \int_{C_n} \frac{f(\zeta)}{(\zeta - z)\sin^2(\pi/h)\zeta} d\zeta, \quad (52)$$

in which the paths  $C_n$  are the same as given in Fig. 2. This leads to the following theorem.

#### Theorem 2

The entire function  $f(z)$  is given by

$$f(z) = \sum_{i=-\infty}^{\infty} \left[ \frac{\sin(\pi/h)(z - jh)}{(\pi/h)(z - jh)} \right]^2 [f_i + (z - jh)f'_i],$$

in which the series is uniformly convergent in any finite closed domain of the  $z$ -plane,  $f_i = f(jh)$ ,

$$f'_i = \left. \frac{df(z)}{dz} \right|_{z=jh}, \quad h = |h| e^{i\varphi},$$

if there is a constant  $K$  so that

$$e^{-(2\pi/|h|)|y|} A(y) \leq \frac{K}{|y|}, \quad \text{for } |y| \rightarrow \infty.$$

Since the proof of this theorem follows the same lines as for theorem 1, it will be omitted.

The interpolation series consists of two convolution sums whose kernels are

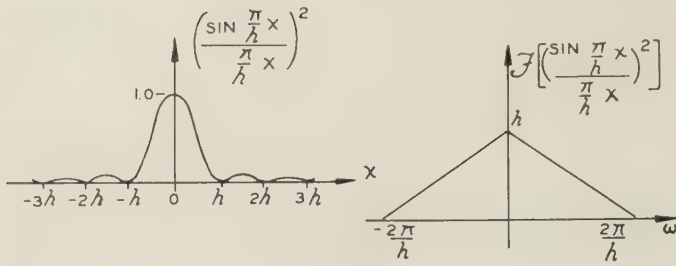
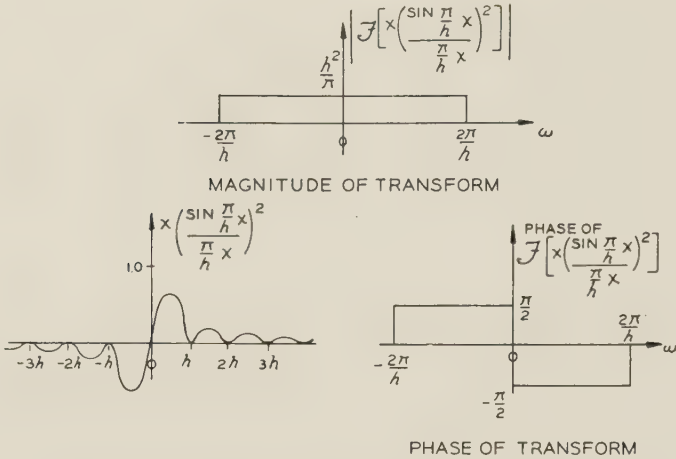
$$\left( \frac{\sin(\pi/h)z}{(\pi/h)z} \right)^2$$

used with the sample values  $f_i$ , and

$$z \left( \frac{\sin(\pi/h)z}{(\pi/h)z} \right)^2$$

used with  $f'_i$ . Graphs of these kernels and their transforms are presented in Figs. 5 and 6.

<sup>5</sup> The approximation is improved when  $\tau$  is decreased.

Fig. 5 The kernel  $(\sin(\pi/h)x/(\pi/h)x)^2$  and its transform.Fig. 6—The kernel  $x(\sin(\pi/h)x/(\pi/h)x)^2$  and its transform.

The realization of this interpolation series is accomplished by using two sequences of pulses whose heights are  $f_i$  and  $f'_i$  with common base width  $\tau$ . As in the case of the cardinal series, when integral convolutions are used the output will closely approximate  $\tau f(z)$  if  $f(z)$  satisfies the conditions of theorem 2.

A class of functions which is of particular interest is the class of band-limited functions defined as follows: A function  $f(z)$  is said to be band-limited if there exists a constant  $W > 0$  and a function  $g^{(\omega)}$  of bounded variation over the interval  $(-2\pi W, 2\pi W)$  so that

$$f(z) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{i\omega z} dg(\omega). \quad (53)$$

The function  $g(\omega)$  is called the Fourier Stieltjes spectrum of  $f(z)$ , and  $W$  the maximum frequency. In case  $g(\omega)$  is absolutely continuous over  $(-2\pi W, 2\pi W)$ , then (53) may be written

$$f(z) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{i\omega z} g'(\omega) d\omega \quad (54)$$

and  $g'(\omega)$  is called simply the Fourier spectrum of  $f(z)$ . Clearly a band-limited function is entire.

From (53)

$$|f(z)| \leq \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{-\omega y} dV(\omega) \quad (55)$$

in which  $V(\omega)$  is the variation of  $g(\omega)$  over  $(-2\pi W, \omega)$ .

Let

$$V = \int_{-2\pi W}^{2\pi W} dV(\omega), \quad (56)$$

that is,  $V$  is the total variation of  $g(\omega)$  over  $(-2\pi W, 2\pi W)$ , then, from (55),

$$|f(z)| \leq \frac{V}{2\pi} e^{2\pi W|y|}. \quad (57)$$

Thus one may define the function  $A(y)$  by

$$A(y) = \frac{V}{2\pi} e^{2\pi W|y|}. \quad (58)$$

Consider the product

$$e^{-(\pi/|h|)|y|} A(y) = \frac{V}{2\pi} e^{\pi(2W-1/|h|)|y|}; \quad (59)$$

clearly the inequality

$$e^{-(\pi/|h|)|y|} A(y) \leq \frac{K}{|y|}, \quad |y| \rightarrow \infty \quad (60)$$

is satisfied if

$$2W - \frac{1}{|h|} < 0. \quad (61)$$

Hence, referring to theorem 1, the following theorem may be enunciated.

### Theorem 3

A band-limited function  $f(z)$  with maximum frequency  $W$  is representable by

$$f(z) = \sum_{j=-\infty}^{\infty} f_j \frac{\sin(\pi/h)(z - jh)}{(\pi/h)(z - jh)}$$

provided the sampling interval  $h$  satisfies

$$|h| < \frac{1}{2W}. \quad (62)$$

Let  $zf(z)$  be band-limited and  $\lim_{z \rightarrow 0} zf(z) = 0$ , then,  $f(z)$  is entire and, from (57),

$$|f(z)| \leq \frac{V}{2\pi} \frac{e^{2\pi W|y|}}{|y|} \quad (63)$$

hence theorem 4.

### Theorem 4

If the function  $zf(z)$  is band-limited with maximum frequency  $W$  and  $\lim_{z \rightarrow 0} zf(z) = 0$ , then  $f(z)$  is representable by the cardinal series provided  $|h| \leq 1/2W$ .

Let  $f(z)$  be band-limited over the interval  $(-2\pi W, 2\pi W)$  with a Fourier spectrum  $g(\omega)$  of bounded variation, then

$$f(z) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{i\omega z} g(\omega) d\omega. \quad (64)$$

Define  $g(\omega)$  to be zero outside the interval  $(-2\pi W, 2\pi W)$ , then

$$f(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega z} g(\omega) d\omega. \quad (65)$$

Integration by parts shows that

$$zf(z) = -\frac{1}{2\pi i} \int_{-2\pi W}^{2\pi W} e^{i\omega z} dg(\omega), \quad (66)$$

that is,  $zf(z)$  is band-limited and from (64)  $\lim_{z \rightarrow 0} zf(z) = 0$ . Hence theorem 5.

#### Theorem 5

If  $f(z)$  is band-limited with maximum frequency  $W$  and Fourier spectrum of bounded variation over the interval  $(-2\pi W, 2\pi W)$ , then  $f(z)$  is representable by the cardinal series provided  $|h| \leq 1/2W$ .

Representation of  $f(z)$  by means of the interpolation series of theorem 2 is immediately obtained from the above analysis. Theorems 6, 7, and 8 follow.

#### Theorem 6

A band-limited function  $f(z)$  with maximum frequency  $W$  is representable by

$$f(z) = \sum_{i=-\infty}^{\infty} \left[ \frac{\sin(\pi/h)(z - jh)}{(\pi/h)(z - jh)} \right]^2 [f_i + (z - jh)f'_i]$$

provided  $|h| < 1/W$ .

#### Theorem 7

If the function  $zf(z)$  is band-limited with maximum frequency  $W$  and  $\lim_{z \rightarrow 0} zf(z) = 0$ , then  $f(z)$  is representable by the interpolation series of theorem 2 provided

$$|h| \leq \frac{1}{W}.$$

#### Theorem 8

If  $f(z)$  is band-limited with maximum frequency  $W$  and Fourier spectrum of bounded variation over the interval  $(-2\pi W, 2\pi W)$ , then  $f(z)$  is representable by the interpolation series of theorem 2 provided  $|h| \leq 1/W$ .

Theorem 2 and the last three theorems indicate that the use of both  $f_i$  and  $f'_i$  allows exact representation using a sample interval twice that required for representation of the same function when using  $f_i$  alone. The above presents a procedure which may be used to develop interpolation series applicable to the general case where  $f_i, f'_i, f''_i, \dots, f_i^{(n)}$  values are available. This case has been considered in a previous paper [9], where it was indicated that, if  $n$  derivatives are available as discrete data, then exact representation can be obtained provided that

$$|h| \leq \frac{n+1}{2W}. \quad (67)$$

It is interesting to note that this result agrees with the representation in  $2WT$  dimensional signal-space, where  $T$  is the time interval duration outside of which the sample values,  $f_i$ , are small. It appears that the number of independent samples or orthogonal dimensions can be arbitrarily chosen from among the continuous function sample values or its derivatives.

#### CONCLUSION

Several applications of this interpolation analysis appear possible in connection with engineering systems; for example, ordinary pulse radar wherein the slant range is determined by the return-time interval has implicit range rate information contained in the Doppler modulation of the pulse carrier frequency. Thus, if slant range is the desired continuous variable, then the available discrete data may be used to establish an approximation of the function and its first derivative.

Obvious additional applications occur in the field of air traffic control wherein the aircraft estimated velocity as well as position is used upon which to base a continuous course plot of the air-path.

The human operator in vehicular control normally uses an interpolation technique wherein the input data is the function and its first derivative at discrete points in time. This is due to the fact that the operator observes each display at discrete times and almost always perceives rate as well as position of a pointer. It then becomes immediately evident that almost all information possible about the amplitude variations corresponding to the pointer position (relating to some parameter) can be discerned by observation of the display at times almost half as often as when the pointer position alone is observed. It is not intended to reify this analysis. However, certain aspects of human engineering may now become more apparent and allow for numerical approximation through the application of the above stated theorems.

#### APPENDIX

Let the sample points be equidistant then the following scale change may be made.

$$t = hw, \quad (68)$$

$$t_i = hw_i. \quad (69)$$

Hence, by substitution,

$$\begin{aligned} L_i^n(w) &= \frac{(hw - hw_0) \cdots (hw - hw_{i-1})(hw - hw_{i+1}) \cdots (hw - hw_n)}{(hw_i - hw_0) \cdots (hw_i - hw_{i-1})(hw_i - hw_{i+1}) \cdots (hw_i - hw_n)} \\ &= \frac{[(w - w_0) \cdots (w - w_{i-1})(w - w_{i+1}) \cdots (w - w_n)]h}{[(w_i - w_0) \cdots (w_i - w_{i-1})(w_i - w_{i+1}) \cdots (w_i - w_n)]h} \quad (70) \end{aligned}$$

Thus, equidistant data may be shifted to the integral time points without loss of generality and no effect on the Lagrangean Coefficients.

## BIBLIOGRAPHY

- [1] Whittaker, J. M., *Interpolatory Function Theory*. ("Cambridge Tracts in Mathematics and Mathematical Physics," No. 33.) London: Cambridge University Press, 1935.
- [2] Steffensen, J. F., *Interpolation*. New York: Chelsea Publishing Co., 1950.
- [3] Nörlund, N. E. "Sur les Formules d'Interpolation de Stirling et Newton," *Annales Science de l'Ecole Normale*, vol. 39 (1922), pp. 343-403; vol. 40 (1923), pp. 35-54.
- [4] Nörlund, N. E. *Les Series d'Interpolation*, Paris: 1926.
- [5] Whittaker, E. T., "On the Functions which are Represented by the Expansions of the Interpolation—Theory," *Proceedings of Royal Society of Edinburgh*, vol. 35 (1915), pp. 181-194.
- [6] Ferrar, W. L., "On the Cardinal Function of Interpolation—Theory," *Proceedings of Royal Society of Edinburgh*, vol. 45 (1925), pp. 269-282; vol. 46 (1925), pp. 323-333.
- [7] Ferrar, W. L., "On the Consistency of Cardinal Function Interpolation," *Proceedings of Royal Society of Edinburgh*, vol. 47 (1927), pp. 230-242.
- [8] Goldman, S. *Information Theory*. New York: Prentice-Hall, Inc., 1953.
- [9] Fogel, L., "A Note on the Sampling Theorem," *IRE TRANSACTIONS*, vol. IT-1 (March, 1955), pp. 47-48.

# The Axis-Crossing Intervals of Random Functions\*

J. A. McFADDEN†

**Summary**—For an arbitrary random process  $\xi(t)$  there exists a function  $x(t)$  which may be obtained by infinite clipping. The axis crossings of  $x(t)$  are identical with those of  $\xi(t)$ . This paper relates the probability density  $P(\tau)$  of axis-crossing intervals to  $r(\tau)$ , the autocorrelation function of  $x(t)$ , i.e., the autocorrelation after clipping. It is shown that the expected number of zeros per unit time is proportional to  $r'(0+)$ , i.e., the right-hand derivative of  $r(\tau)$  at  $\tau = 0$ . Next a theorem is proved, stating that  $P(\tau) = 0$  over a finite range  $0 \leq \tau < T$  if and only if  $r(\tau)$  is linear in  $|\tau|$  over the corresponding range of  $|\tau|$ . If  $r(\tau)$  is nearly linear for small  $\tau$ , then the initial behavior of  $P(\tau)$  is like  $r''(\tau)$ . These results are illustrated by some simple, random square-wave models and by a comparison with Rice's results for Gaussian noise.

## INTRODUCTION

ONE OF THE outstanding unsolved problems in the mathematical theory of noise is the determination of the distribution of intervals between axis crossings. For Gaussian noise, Rice<sup>1</sup> has given an approximate result and has discussed the difficulties which impede a more accurate solution. Little is known, however, about the interval distribution for non-Gaussian noise. The present paper provides a new approach to the problem, whereby some of Rice's results may be generalized for noise with an arbitrary distribution of instantaneous amplitudes.

When a noise signal is passed through an infinite clipper, the axis crossings are invariant but all other information is lost. In this paper we attempt to relate the axis-crossing interval distribution to the properties of the signal after clipping.

The signal after clipping is defined as follows: Let  $x(t)$  describe a random process which is both stationary and ergodic.  $x(t)$  may assume only the values  $\pm 1$ , and

either value is equally likely. Let the autocorrelation function for this process, i.e., the expectation  $E[x(t)x(t + \tau)]$ , be denoted by  $r(\tau)$ .

We define the probability density  $P(\tau)$  for axis-crossing intervals in the following manner: Given a zero at time  $t$ ,  $P(\tau)d\tau$  is the conditional probability that the next zero lies between  $t + \tau$  and  $t + \tau + d\tau$ . The integral of  $P(\tau)d\tau$  from 0 to  $\infty$  is unity, and the mean axis-crossing interval is

$$E(\tau) = \int_0^\infty \tau P(\tau) d\tau. \quad (1)$$

Let  $\beta$  be the expected number of zeros per unit time. Then  $E(\tau)$  is the reciprocal of  $\beta$ , for in a very long time interval  $K$  there will be  $\beta K$  zeros and  $\beta K$  intervals. The mean interval is the sum of all the intervals,  $K$ , divided by the number of intervals, and the result is  $1/\beta$ .

The purpose of this paper is to relate the interval density  $P(\tau)$  to the correlation  $r(\tau)$  for an arbitrary function of the type  $x(t)$ .

## EXAMPLES OF $x(t)$

We shall cite four examples of the process  $x(t)$ . These will be used later to check the general results. The first three are defined without reference to the clipping process.

### Periodic Square Wave

The first example of  $x(t)$  is the periodic square wave with unit amplitude, period  $2T$ , and random time origin, distributed uniformly between the values 0 and  $2T$ .

The correlation function  $r(\tau)$  is a triangular wave,

$$r(\tau) = 1 - \frac{2|\tau|}{T}, \quad -T < \tau < T, \quad (2)$$

$$r(\tau + 2T) = r(\tau).$$

\* Manuscript received by the PGIT, May 25, 1956.

† U. S. Naval Ordnance Lab., Silver Spring, Md.

<sup>1</sup> S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282-332; July, 1944; vol. 24, pp. 46-156; January, 1945; see sec. 3.4.

In this case the solution for  $P(\tau)$  is trivial, since all the axis-crossing intervals have a length  $T$ . We have, using the Dirac  $\delta$  function,

$$P(\tau) = \delta(\tau - T). \quad (3)$$

#### Random Sequence of Pulses<sup>2</sup>

The second example is a random sequence of pulses of duration  $T$ . At an initial instant  $t_0$  we toss a symmetric coin to determine whether  $x(t_0)$  shall be  $+1$  or  $-1$ . Whatever the result,  $x(t)$  remains constant until  $t = t_0 + T$ ; then the coin is tossed again, etc.,  $t_0$  is a random variable, distributed uniformly between 0 and  $T$ .

The correlation function consists of a single triangular peak.<sup>2</sup>

$$r(\tau) = 1 - \frac{|\tau|}{T}, \quad 0 \leq |\tau| < T; \quad (4)$$

$$= 0, \quad |\tau| \geq T.$$

The solution for the axis-crossing interval density can be determined easily. The only intervals possible are multiples of  $T$ . The probability that an interval is  $(n+1)T$  is one-half the probability that it is  $nT$ . It follows directly that

$$P(\tau) = \sum_{n=1}^{\infty} 2^{-n} \delta(\tau - nT). \quad (5)$$

#### Markoff Process

The third example of  $x(t)$  is a Markoff process. In this process, the probability that  $x(t+dt)$  is  $+1$  or  $-1$  depends only on the value of  $x(t)$ . In other words, the probability that  $x(t)$  changes sign between  $t$  and  $t+dt$  is independent of what happens outside this interval. Let this last probability be  $\alpha dt$ , where  $\alpha$  is a constant. Then it is known<sup>3</sup> that the correlation function is

$$r(\tau) = e^{-2\alpha|\tau|}. \quad (6)$$

Furthermore the number of zeros in a given time interval obeys the well-known Poisson distribution,<sup>3</sup> and the probability density of intervals between successive zeros is<sup>4</sup>

$$P(\tau) = \alpha e^{-\alpha\tau}. \quad (7)$$

[This process should not be confused with the Gaussian Markoff (or Ornstein-Uhlenbeck) process, since  $x(t)$  is not Gaussian. If a Gaussian Markoff process is infinitely clipped, the output  $x(t)$  will not be Markoffian.]

#### Clipped Gaussian Noise

The fourth example of  $x(t)$  is infinitely clipped Gaussian noise. Let  $\xi(t)$  describe a Gaussian process, with normalized

autocorrelation function  $\rho(\tau)$ , and let  $x(t)$  be the output voltage after  $\xi(t)$  is infinitely clipped. We have

$$x(t) = \begin{cases} 1 & \text{if } \xi(t) \geq 0; \\ -1 & \text{if } \xi(t) < 0. \end{cases} \quad (8)$$

Then the autocorrelation function of  $x(t)$ , which we call  $r(\tau)$ , is related to  $\rho(\tau)$  by the well known arcsine law,<sup>5</sup>

$$r(\tau) = \frac{2}{\pi} \sin^{-1} \rho(\tau). \quad (9)$$

If  $\rho(\tau)$  is regular at  $\tau = 0$ , then it can be expanded in a power series containing only the terms of even order in  $\tau$ , since  $\rho(\tau)$  is always an even function of  $\tau$ . Then it follows from (9) that  $r(\tau)$  can be expanded in a series containing only odd-order terms in  $|\tau|$ , except for the constant term, which is unity. If  $\rho(\tau)$  is regular, then it is smooth at  $\tau = 0$  and has zero slope there, but  $r(\tau)$  displays a triangular peak with zero curvature in the neighborhood of  $\tau = 0$ . For small  $|\tau|$ , the derivative  $\rho'(\tau)$  is large and negative, but  $r'(\tau)$  ( $\tau \neq 0$ ) is positive and small in magnitude compared to  $\rho''(0)$ . [An exceptional case occurs when  $\xi(t)$  is nondifferentiable; i.e., the derivative  $\xi'(t)$  has infinite power. Take, for example, the Gaussian Markoff (or Ornstein-Uhlenbeck) process, for which  $\rho(\tau)$  is an exponential function<sup>6</sup> of  $|\tau|$ . In this case  $\rho(\tau)$  is not regular, since it displays a triangular peak at  $\tau = 0$ . It follows from (9) that the corresponding  $r(\tau)$  possesses an infinite cusp.]

For Gaussian noise the exact axis-crossing interval distribution has not been determined. The most important work has been that of Rice, who calculated the conditional probability that a downward crossing lies between  $t+\tau$  and  $t+\tau+d\tau$ , given an upward crossing at time  $t$ . Let us call this conditional probability  $Q(\tau)d\tau$ <sup>7</sup>. Since the zero in  $d\tau$  need not be the next zero after time  $t$ , Rice's  $Q(\tau)$  will agree closely with  $P(\tau)$  only when  $\tau$  is small. However, for a narrow-band spectrum, the range of agreement includes most of the practical range of axis-crossing intervals.

#### EXPECTED NUMBER OF ZEROS PER UNIT TIME

We shall now return to the general process  $x(t)$  which we defined earlier.

If  $x_1 = x(t_1)$  and  $x_2 = x(t_2)$ , the probability that  $x_1 = -1$  and  $x_2 = +1$  is  $\frac{1}{4}(1 - r_{12})$ , where  $r_{12}$  is the correlation coefficient between  $x_1$  and  $x_2$ . Let us derive this result. We let the probability that  $x_1$  and  $x_2$  are both equal to  $+1$  be  $P_{++}$ , and similarly we define  $P_{--}$ ,  $P_{+-}$ , and  $P_{-+}$ . Then the moments  $E(1)$ ,  $E(x_1)$ ,  $E(x_2)$ , and  $E(x_1 x_2)$  are obtained as follows:

<sup>2</sup> *Ibid.*, sec. 2.7.  
H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of Servomechanisms," McGraw-Hill Book Co., Inc., New York, N. Y., p. 267; 1947.

<sup>3</sup> Rice, *loc. cit.*, sec. 2.7.

<sup>4</sup> Rice, *loc. cit.*, sec. 3.4, following eq. (3.4-11).

<sup>5</sup> J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," McGraw-Hill Book Co., Inc., New York, N. Y., p. 58; 1950.

<sup>6</sup> J. L. Doob, "The Brownian movement and stochastic equations," *Ann. Math.*, vol. 43, pp. 351-369; April, 1942; see p. 353.

<sup>7</sup> Rice, *loc. cit.*, eq. (3.4-1).

$$\begin{aligned}
E(1) &= P_{++} + P_{--} + P_{+-} + P_{-+} \\
E(x_1) &= P_{++} - P_{--} + P_{+-} - P_{-+} \\
E(x_2) &= P_{++} - P_{--} - P_{+-} + P_{-+} \\
E(x_1x_2) &= P_{++} + P_{--} - P_{+-} - P_{-+}.
\end{aligned} \tag{10}$$

Now  $E(1) = 1$ ,  $E(x_1) = E(x_2) = 0$ , and  $E(x_1x_2) = r_{12}$ ; therefore the solution of these four equations yields the result,  $P_{-+} = \frac{1}{4}(1 - r_{12})$ . Since  $P_{+-} = P_{-+}$ , the probability that  $x_1$  and  $x_2$  are of opposite sign is  $\frac{1}{2}(1 - r_{12})$ .

Now let  $t_1 = t$  and  $t_2 = t + \Delta t$ , where  $\Delta t$  is small enough that the probability of finding more than one zero between  $t$  and  $t + \Delta t$  can be neglected. Then the probability that  $x(t_1)$  and  $x(t_2)$  are of opposite sign becomes the expected number of zeros per unit time multiplied by  $\Delta t$ . Thus

$$\beta \Delta t = \frac{1}{2}[1 - r(\Delta t)]. \tag{11}$$

For sufficiently small  $\Delta t$ ,  $r(\Delta t) = 1 + r'(0+) \Delta t$ ; therefore (11) yields the result,

$$\beta = -\frac{1}{2}r'(0+). \tag{12}$$

$r'(0+)$  denotes the right-hand derivative of  $r(\tau)$  with respect to  $\tau$ , evaluated at  $\tau = 0$ . We must distinguish between  $r'(0+)$  and  $r'(0-)$  because they are never equal.  $r(\tau)$  is an even function; therefore  $r'(\tau)$  is odd. Then  $r'(0+) = -r'(0-)$ , and consequently the two derivatives cannot be equal unless they are zero. They cannot be zero, for then (12) would imply that  $\beta = 0$ , i.e., no axis crossings at all.

$E(\tau)$ , the mean axis-crossing interval, is the reciprocal of this  $\beta$ .

Let us apply (12) to the examples given previously. For the periodic square wave, we may differentiate (2). We have  $r'(0+) = -2/T$ ; therefore  $\beta = 1/T$  and  $E(\tau) = T$ , obviously in agreement with the probability density (3).

In the random sequence of pulses, (4) gives us  $\beta = 1/2T$ ,  $E(\tau) = 2T$ , and this result may be confirmed from the probability density (5).

In the Markoff process, we find from (6) that  $\beta = \alpha$ , in agreement with the definition of  $\alpha$ , and that  $E(\tau) = 1/\alpha$ .

For clipped Gaussian noise, we differentiate (9) with respect to  $\tau$ ; then by a limiting process we find that

$$\beta = \frac{1}{\pi} [-\rho''(0)]^{1/2}, \tag{13}$$

provided  $\rho'(0) = 0$ . This is Rice's result.<sup>8</sup> If  $\xi(t)$  is non-differentiable, then  $\rho'(0) \neq 0$  and  $\beta$  is infinite [for example, if  $\xi(t)$  is the output of an RC low-pass filter and the input is broad-band noise—Rice<sup>8</sup> has discussed this example].

#### A THEOREM CONCERNING $P(\tau)$

Let us now prove a general theorem which relates the axis-crossing interval density  $P(\tau)$  to the autocorrelation  $r(\tau)$ .

#### Theorem 1

$P(\tau) = 0$  over a finite range  $0 \leq \tau < T$ , if and only if  $r(\tau)$  is linear in  $|\tau|$  over the range  $0 \leq |\tau| < T$ . In other words, if the autocorrelation of  $x(t)$  is a linear function of  $|\tau|$  between  $\tau = 0$  and  $|\tau| = T$ , then the probability of an axis-crossing interval shorter than  $T$  is zero, and conversely.

Proof of sufficient condition: The probability that  $x(t_1)$ ,  $x(t_2)$ , and  $x(t_3)$  are respectively  $+1$ ,  $-1$ , and  $+1$  is

$$P_{+-+} = \frac{1}{8}(1 - r_{12} + r_{13} - r_{23}), \tag{14}$$

where the  $r_{ij}$  are the respective correlation coefficients. This result may be derived by a generalization of (10). We write the moments  $E(1)$ ,  $E(x_1)$ ,  $E(x_2)$ ,  $E(x_3)$ ,  $E(x_1x_2)$ ,  $E(x_1x_3)$ ,  $E(x_2x_3)$ , and  $E(x_1x_2x_3)$  in terms of the probabilities  $P_{+++}$ ,  $P_{---}$ , etc. Then we set  $E(1) = 1$ ,  $E(x_i) = 0$ ,  $E(x_ix_j) = r_{ij}$ , and  $E(x_1x_2x_3) = 0$ , and solve for  $P_{+-+}$ .<sup>9</sup>

Now let  $t_1 = t$ ,  $t_2 = t + \tau_1$ , and  $t_3 = t + \tau_2$ . Then (14) becomes

$$P_{+-+} = \frac{1}{8}[1 - r(\tau_1) + r(\tau_2) - r(\tau_2 - \tau_1)]. \tag{15}$$

By the assumption in theorem 1, we may define  $r(\tau)$  as follows:

$$r(\tau) = 1 - a|\tau|, \quad 0 \leq |\tau| < T, \tag{16}$$

where  $a$  is a constant,  $0 < a \leq 2/T$ . Under this assumption, (15) becomes identically zero for all values of  $\tau_1$  and  $\tau_2$  in the range  $0 \leq \tau < T$ . By symmetry the probability  $P_{-+-}$  is also zero. Then we conclude that the probability of finding more than one zero in an interval of length  $\tau$  is zero, when  $\tau < T$ ; therefore we have proved sufficiency in theorem 1.

A different procedure will be employed for the proof of the necessary condition.

Proof of necessary condition: The autocorrelation  $r(\tau)$  may be written as a time average,

$$r(\tau) = \lim_{L \rightarrow \infty} \frac{1}{2L} \int_{-L}^L x(t)x(t+\tau) dt, \tag{17}$$

where  $L$  is a very long period of time. The derivative  $r'(\tau)$  may be written as follows:

$$r'(\tau) = \lim_{L \rightarrow \infty} \frac{1}{2L} \int_{-L}^L x(t)x'(t+\tau) dt. \tag{18}$$

Now  $x(t)$  is always  $+1$  or  $-1$ ; therefore  $x'(t)$  is a sequence of positive and negative spikes or  $\delta$  functions,

$$x'(t) = 2 \sum_i (-1)^i \delta(t - t_i), \tag{19}$$

where  $t_i$  are the successive zeros of  $x(t)$ . The factor 2 arises because  $x(t)$  changes by  $\pm 2$  at each crossing.

The integrand of (18) will be zero except when  $t + \tau = t_i$ . Let  $\tau$  be in the range  $0 < \tau < T$ . If  $t_i$  is an upward crossing, then  $x(t_i - \tau)$  will be negative, provided there are no

<sup>9</sup> For a generalization to  $n$  variables, see J. A. McFadden "Urn models of correlation and a comparison with the multivariate normal integral," *Ann. Math. Statist.*, vol. 26, pp. 478-489; September, 1955, sec. 6.

<sup>8</sup> Rice, *loc. cit.*, eq. (3.3-11).

axis-crossing intervals shorter than  $\tau$ . This last condition is fulfilled by the assumption in theorem 1. On the other hand, if  $t_i$  is a downward crossing, then  $x(t_i - \tau)$  will be positive. In either case, the product  $x(t_i - \tau)x'(t_i)$  will be negative. For any long interval  $(-L, L)$  the integral will be twice the number of zeros in the interval, with a negative sign. The number of zeros is  $2L$  times the expected number per unit time, or  $2\beta L$ ; therefore (18) yields

$$r'(\tau) = -2\beta, \quad 0 < \tau < T. \quad (20)$$

By a similar argument for negative  $\tau$ , we have

$$r'(\tau) = 2\beta, \quad -T < \tau < 0. \quad (21)$$

At the point  $\tau = 0$ ,  $r'(\tau)$  is discontinuous. But  $r'(0)$  is bounded, since it must lie between  $-2\beta$  and  $+2\beta$ . Then, by integration of the above equations, together with the initial condition on the correlation function,  $r(0) = 1$ , we find that

$$r(\tau) = 1 - 2\beta |\tau|, \quad 0 \leq |\tau| < T. \quad (22)$$

Since  $\beta$  is a constant, we have proved the second half of theorem 1. [Note: If we let  $\tau \rightarrow 0$  from the positive side in (20), the result checks (12).]

Let us examine the four previous examples of  $x(t)$ . The periodic square wave with period  $2T$  and the random sequence of pulses of width  $T$  are in accord with theorem 1, and the definitions of  $T$  are consistent.

The Markoff process has no linear range for  $r(\tau)$ ; in fact, the second derivative  $r''(\tau)$  is positive at  $\tau = 0 +$  and at  $\tau = 0 -$ . There is no finite range  $0 \leq \tau < T$  for which  $P(\tau) = 0$ ; even for small  $\tau$ ,  $P(\tau)$  is not a small quantity.

Clipped Gaussian noise has no finite linear range for  $r(\tau)$ , but if  $\rho(\tau)$  is regular the curvature of  $r(\tau)$  will be small for small  $|\tau| > 0$ , since  $r''(0+) = r''(0-) = 0$ . Also, from Rice's work<sup>1</sup> we know that  $P(\tau)$  is small for small  $\tau$ . Thus our theorem is applicable only in a qualitative, intuitive sense. If the Gaussian noise has an extremely narrow bandwidth, then  $\rho(\tau)$  is nearly sinusoidal and  $r(\tau)$  is nearly the triangular wave (2). As the bandwidth (before clipping) approaches zero, clipped Gaussian noise approaches the periodic square wave, and the conditions in our theorem become more nearly applicable.

#### DERIVATION OF $P(\tau)$ FOR SMALL $\tau$

We are now ready to derive an approximation for the axis-crossing interval density  $P(\tau)$  when the intervals are small, for an arbitrary random function of the type  $x(t)$ .

We have already seen that the existence of a nonzero  $P(\tau)$  is somehow associated with the curvature of the correlation function  $r(\tau)$ . We shall now derive the result that, for small  $\tau$ ,  $P(\tau)$  is directly proportional to  $r''(\tau)$ .

Assumption: We assume that there exists a quantity  $T_1$  such that the probability of more than two zeros occurring in an interval  $(t, t + T_1)$  is so small that it may be neglected. This condition can usually be fulfilled by taking  $T_1$  small enough; the choice of  $T_1$  depends on the particular model

and on the accuracy desired. For example, if most of the axis-crossing intervals lie near the mean interval  $E(\tau)$ , then we may choose  $T_1$  slightly smaller than  $2E(\tau)$ , since the occurrence of two successive axis-crossing intervals having a sum less than this  $T_1$  will be rare.

Under this assumption we may proceed as in the proof of the second half of theorem 1. We shall apply (18) when  $\tau$  is in the range  $0 < \tau < T_1$ . Suppose  $t_i$  is an upward crossing in  $x(t)$ , and let  $u_i$  be the preceding axis-crossing interval; i.e.,  $u_i = t_i - t_{i-1}$ . Then  $x(t_i - \tau)$  will be negative if  $\tau < u_i$  and positive if  $\tau > u_i$ .  $t_i - \tau$  cannot be located before the next preceding zero  $t_{i-2}$ , because  $\tau < T_1$  and because of our basic assumption. Now the probability that  $u_i > \tau$  or  $u_i < \tau$  is

$$\int_{\tau}^{\infty} P(u) du \quad \text{or} \quad \int_0^{\tau} P(u) du,$$

respectively. Then from (18) we find

$$r'(\tau) = -2\beta \int_{\tau}^{\infty} P(u) du + 2\beta \int_0^{\tau} P(u) du. \quad (23)$$

But the integral from 0 to  $\infty$  is unity; therefore we have

$$r'(\tau) = -2\beta + 4\beta \int_0^{\tau} P(u) du. \quad (24)$$

Now let us solve for  $P(\tau)$ . If  $r''(\tau)$  exists, we may differentiate (24). Then if we eliminate  $\beta$  by (12), we obtain the result,

$$P(\tau) = \frac{r''(\tau)}{-2r'(0+)}. \quad (25)$$

If  $r'(\tau)$  is discontinuous at  $\tau = \tau_1$ , then the density  $P(\tau_1)$  is a  $\delta$  function. We may calculate the probability that  $\tau = \tau_1$  by evaluating (24) at  $\tau_1 + 0$  and at  $\tau_1 - 0$ , and then taking the difference. Then

$$Pr\{\tau = \tau_1\} = \frac{r'(\tau_1+) - r'(\tau_1-)}{-2r'(0+)}. \quad (26)$$

Eqs. (25) and (26) are the desired results for the range  $0 < \tau < T_1$  and under the given assumption. Let us return to our examples of  $x(t)$ .

The periodic rectangular wave is trivial. The above formulas yield  $Pr\{\tau = T\} = 1$  and  $P(\tau) = 0$  elsewhere. This result agrees exactly with (3), since the assumption at the beginning of this section is correct when  $T_1 < 2T$ .

For the random sequence of pulses, we find  $Pr\{\tau = T\} = 1/2$  and  $P(\tau) = 0$  elsewhere in the range  $0 < \tau < 2T$ . This result agrees with (5), but we cannot obtain any further information about  $P(\tau)$ . The basic assumption, not more than two zeros in an interval of width  $T_1$ , is fulfilled when  $T_1 < 2T$  but not when  $T_1 \geq 2T$ .

The example in which  $x(t)$  is a Markoff process is not at all suited to the above theory. Since the number of zeros in a given interval obeys the Poisson distribution, there is no finite  $T_1$  for which the basic assumption holds even approximately.

For clipped Gaussian noise, we may substitute the arcsine law (9) into (25). Then we have

$$P(\tau) = \frac{1}{2\pi\beta} \frac{\rho(\tau)[\rho'(\tau)]^2 + [1 - \rho^2(\tau)]\rho''(\tau)}{[1 - \rho^2(\tau)]^{3/2}}, \quad (27)$$

where  $\beta$  is given by (13). This is identical with an approximation given by Rice for the function we have called  $Q(\tau)$ .

Let us review briefly the method of Rice. Instead of calculating  $P(\tau)$ , Rice calculated  $Q(\tau)$ , which we have already defined.  $Q(\tau)$  is always a good approximation to  $P(\tau)$  if  $\tau$  is small enough. However, if  $\xi(t)$  has a narrow-band spectrum, then most of the axis-crossing intervals  $\tau$  are near the mean value  $E(\tau)$ . For intervals from 0 up to a value slightly below  $3E(\tau)$ ,  $Q(\tau)$  will be very close to  $P(\tau)$ , since for this range a downward crossing between  $t + \tau$  and  $t + \tau + d\tau$  is very likely the *next* crossing. In other words,  $Q(\tau)$  always gives the initial behavior of  $P(\tau)$ , but in the narrow-band case the range of validity extends upward to include most of the practical range of  $\tau$ .

Rice's expression<sup>7</sup> for  $Q(\tau)$  is extremely complicated. For this reason, Rice specialized to the narrow-band case. For  $\tau$  in the neighborhood of the first maximum of  $Q(\tau)$ , he obtained the approximation (27). He then simplified (27) even further by additional approximations.

In deriving (27), we have not assumed a narrow bandwidth, but only that  $\tau < T_1$ . However, if the noise (before clipping) does have a narrow bandwidth, then  $T_1$  can be chosen slightly smaller than  $2E(\tau)$ , since most of the axis-crossing intervals are near the mean. Then the range of validity of (27),  $0 \leq \tau < T_1$ , will include much of the practical range of  $\tau$ . By our derivation, as contrasted with Rice's, we conclude that the approximation (27) (in the narrow-band case) is valid not only near the peak of  $P(\tau)$  but also for all  $\tau$  to the left of it.

In other words, a narrow bandwidth is not essential to the derivation of (27) for *small*  $\tau$ , but it is essential if (27) is to provide a practical approximation over the important range of  $\tau$ . For example, for the ideal low-pass filter as plotted by Rice,<sup>1</sup> (27) falls about five per cent below Rice's  $Q(\tau)$  at the first maximum of  $Q(\tau)$  but fits  $Q(\tau)$  increasingly better as  $\tau$  becomes smaller.  $Q(\tau)$  itself is

reliable [as an approximation to  $P(\tau)$ ] slightly beyond the peak and everywhere to the left. For  $\tau$  beyond the peak, (27) should not be used for the low-pass filter.

## DISCUSSION

Let us return to the general problem concerning the random function  $x(t)$ . The main purpose of this paper is not to rederive special results already known, but to discuss the relations between the axis-crossing interval density  $P(\tau)$  and the autocorrelation function  $r(\tau)$ . Now that we have checked our general relations with existing results, let us consider a more general noise process, stationary and ergodic (but not necessarily Gaussian), described by the function  $\xi(t)$ . Let the output after infinite clipping be  $x(t)$ , according to (8). The previous analysis indicates that  $P(\tau)$  can be related to the statistical properties of  $x(t)$  much more readily than to those of  $\xi(t)$ . In general however, the statistical properties of  $x(t)$  [e.g., the arcsine relation (9)] will not be known.

What can we say about  $P(\tau)$  if we do know  $r(\tau)$ , but have no further information?

1) In any case, the mean value  $E(\tau)$  is the reciprocal of  $\beta$ , where  $\beta$  is given by (12).

2) If  $r(\tau)$  is nearly linear over a substantial range  $0 < \tau < T$ , then the number of axis-crossing intervals in this range will be very small, and therefore we may choose  $T_1 = 2T$ . The basic assumption in the previous section is approximately fulfilled; therefore (25) provides an approximation to  $P(\tau)$  in the range  $0 < \tau < T_1$ .

We have already mentioned the difficulties inherent in transforming the present theory into practical computations. We feel, however, that the results of this paper provide a useful background for further research which is now in progress, and we hope that they may be of assistance to others as well.

## ACKNOWLEDGMENT

The author is indebted to Gilbert Lieberman for much helpful criticism and discussion on this problem.



# Determination of Redundancies in a Set of Patterns\*

ARTHUR GLOVAZKY†

**Summary**—A set of black-and-white patterns can be identified by successive sampling of the individual “cells” which constitute these patterns. If the number of patterns is  $P$  and the number of cells is  $C$ , it is possible to find, for any specified sampling sequence at least  $C - P + 1$  cells which may be omitted without obstructing unique identification.

Such redundant cells can be found by two methods: Construction of a “code mobile,” and compilation of a “code schedule.” The “mobile” is useful inasmuch as its topological characteristics can be correlated with the information capabilities and the inherent redundancy of the given identification process. The “schedule,” on the other hand, is the numerical means by which practical cases can be rapidly and systematically solved.

Besides revealing the redundancies in a given set, both the “mobile” and the “schedule” may serve as useful tools in evaluating and designing sampling programs.

## INTRODUCTION

A “SET OF PATTERNS,” for the purpose of this paper, is any group of geometrical configurations which obey the following conditions: 1) The number of patterns  $P$  is finite; 2) all the patterns are of the “two-tone” (black-and-white) variety; and 3) any of the given patterns can be divided into a finite number of cells  $C$ , each of which is either wholly black or wholly white.

If these restrictions are met, unique identification of any pattern in the given set can be accomplished by successive sampling of the  $C$  cells. Assuming any particular sampling sequence or *scanning path*, it is conceivable that identification can still be carried out even though some of the cells are by-passed. The objective of the following study is to devise a method by which the cells that are redundant with respect to the given identification process can be determined.

## THE CODE MOBILE

To gain a better insight into the flow and accumulation of information involved in pattern recognition, it is useful to have a close look at the so called *code mobile*.

If the pattern set and the scanning path are known, one can readily construct such a “mobile” by sketching, for each pattern, a chain of “black” and “white” links which corresponds to the sequence of black and white cells in the scanning path. In drawing the chains it is convenient to divert the black links consistently to the left, and the white links consistently to the right. If all the chains are permitted to originate from the same point, some of the links corresponding to various chains will merge and the result would be a structure composed of a number of interconnected *subchains*. The points of interconnection, the *nodes* of the mobile, are quite significant inasmuch as they are the mainsprings, so to speak, of the

whole structure. By construction, these nodes place in evidence the differences between the various patterns and indicate at which stages of the sampling process such differences reveal themselves.

Fig. 1 presents, as a simple example, an arbitrary set of patterns with  $C = 9$ . Fig. 2 indicates, by assigning numbers to the various cells, a possible scanning path. The resulting mobile is shown in Fig. 3, where the numbers at the bottom and on the right refer to patterns and cells respectively.

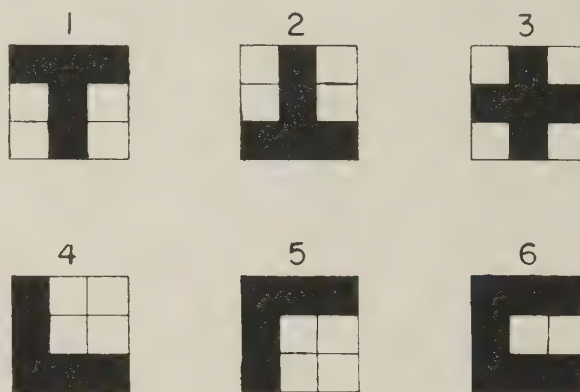


Fig. 1—An arbitrary set of 9 patterns.

1	2	3
4	5	6
7	8	9

Fig. 2—A possible scanning path.

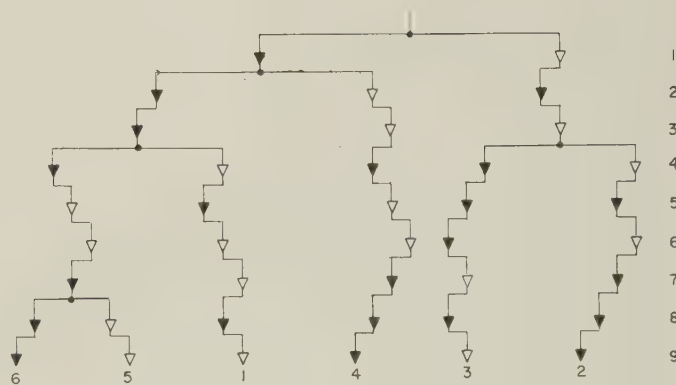


Fig. 3—A code mobile pertinent to Figs. 1 and 2.

It is seen that a code mobile, in our application, is composed of a number of *levels* which equals the number of cells  $C$ . By advancing from one level to the next we can

\* Manuscript received by the PGIT, April 30, 1956.

† Raytheon Mfg. Co., Waltham, Mass.

observe how the  $P$  given patterns are separated into smaller and smaller groups until, finally, each one of them is completely isolated. This process of successive separation, placed in evidence by the nodes, is the essence of the identification process and the only useful operation as far as this process is concerned. It follows, therefore, that any level which does not contain any nodes is redundant and may be removed. As a verification it may be noticed that removal of such levels does not affect the original number of subchains and therefore the original number of distinct patterns which the mobile represents. In terms of the physical recognition mechanism, this result means that the cells which correspond to such redundant levels may be omitted from the scanning path without hindering the unique identification of all the given patterns.

In the mobile of Fig. 3 the levels which do not contain any nodes are 3, 5, 6, 7, and 9. Fig. 4 shows the "reduced" mobile of Fig. 3 after the redundant levels have been removed. It is evident that the reduced mobile still represents 6 distinct patterns, although the number of sampled cells has been reduced from 9 to 4.

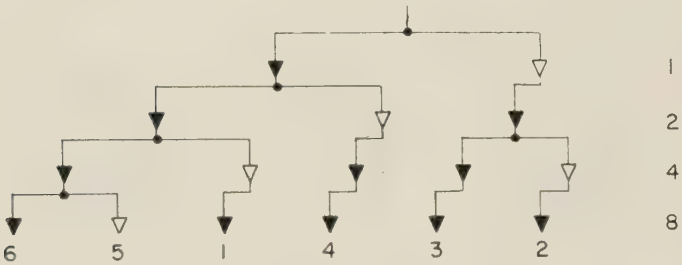


Fig. 4—The reduced mobile of Fig. 3.

THE CODE SCHEDULE

Although the code mobile is very useful as a pictorial aid in understanding the principles involved in the recognition process, in most practical problems its construction may become quite cumbersome. A typical example is the determination of redundancies in the alpha-numeric characters where  $P = 35$  and  $C = 100$ . The impracticability of the mobile in such a case is apparent.

A considerable amount of compactness and facility is gained if the chains associated with the various patterns are expressed in binary digits, where "0" and "1" represent the white and the black cells respectively. The result is a series of  $P$  binary numbers, each containing  $C$  digits. Table I represents the "digitization" of Fig. 1 with respect to the scanning path of Fig. 2; the numbers on the left and at the top of the table refer to patterns and cells respectively.

Let us now arrange the  $P$  numbers in the same sequence that their respective chains appear—from left to right—in the code mobile. It is observed that, thus arranged, every number will be numerically smaller than the preceding one. This follows from the fact that, in con-

structing the mobile, all the "1" links were continuously driven to the left, while all the "0" links were continuously driven to the right; as a result, every chain represents a number which is smaller than the number represented by the chain to its left.

Once the  $P$  numbers are arranged in a sequence of descending numerical value, they faithfully represent the corresponding mobile and may provide all the information previously derived from it. It is convenient to set these numbers in a matrix-like array, called a *code schedule*. This schedule would contain  $P$  rows and  $C$  columns and its  $mn$ 'th element would describe the  $n$ 'th cell of the  $m$ 'th pattern.

From previous discussion it is evident that if, by some method, we could locate the nodes of the mobile on the corresponding schedule, the redundant cells could immediately be determined. If we now recall the property of nodes to effect successive separation, and note that each separation occurs whenever a pair of black and white links emerges from a subchain, we can readily formulate the required procedure.

Proceeding from left to right, the initial step is to find the first column in which 1 neighbors 0 and draw a *separation line* between the two digits which extends from the column in question to the last column. The arrays on both sides of this line may now be regarded as new schedules and the same step is applied to each one of them separately. This procedure is repeated until all the columns are exhausted or until each one of the resulting "subschedules" contains only one row. The nodes of the mobile are now represented on the schedule by the *starting points of the various separation lines*. The columns which do not contain any such starting points correspond, therefore, to the redundant cells.

A better understanding of the procedure may perhaps be achieved through a specific example. Arranging the rows of Table I in a sequence of descending numerical

TABLE I  
DIGITAL REPRESENTATION PERTINENT TO FIGS. 1 AND 2

	1	2	3	4	5	6	7	8	9
1	1	1	1	0	1	0	0	1	0
2	0	1	1	0	1	0	1	1	1
3	0	1	0	1	1	1	0	1	0
4	1	0	0	1	0	0	1	1	1
5	1	1	1	1	0	0	1	0	0
6	1	1	1	1	0	0	1	1	1

value results in the code schedule of Table II. Column 1 of the schedule is the first column to contain a 1-0 pair and therefore to initiate a separation line. From column 2 on, two separate schedules have to be considered. In column 2 one of these "subschedules" contains a 1-0 pair and a second separation line is initiated. From column 3 on, therefore, three schedules have to be considered. In column 4 two of the three subschedules initiate separation

TABLE II

THE CODE SCHEDULE FOR TABLE I AND THE ASSOCIATED SEPARATION LINES

	1	2	3	4	5	6	7	8	9
6	1	1	1	1	0	0	1	1	1
5	1	1	1	1	0	0	1	0	0
1	1	1	1	0	1	0	0	1	0
4	1	0	0	1	0	0	1	1	1
3	0	1	0	1	1	1	0	1	0
2	0	1	0	0	1	0	1	1	1

lines, while in columns 5, 6, and 7 no lines are initiated since none of the arrays in question contains a 1-0 pair. Row 8 initiates the fifth and the last line. It is now evident that columns 3, 5, 6, 7, and 9, and therefore the corresponding cells, are redundant. The "reduced" schedule, composed of columns 1, 2, 4 and 8, is shown in Table III.

TABLE III

THE REDUCED SCHEDULE OF TABLE II

	1	2	4	8
6	1	1	1	1
5	1	1	1	0
1	1	1	0	1
4	1	0	1	1
3	0	1	1	1
2	0	1	0	1

The validity of the reduction is verified by noticing that the rows in the reduced schedule are numerically distinct, and that if any of the remaining columns is removed, this distinctness is destroyed.

As a summary, we shall repeat the steps necessary for the determination of redundant cells associated with any specified set and scanning path: 1) Digitize the patterns at the given sampling sequence; 2) form the code schedule by arranging the resulting numbers in a sequence of descending numerical value; 3) draw the separation lines, and 4) observe the columns which do not initiate any separation lines. The cells corresponding to these columns are the redundant cells.

#### MORE ABOUT SCANNING PATHS

It should be emphasized that the reduced number of cells, as determined above, is the minimum permissible number for the *given scanning path*. It is quite possible

that a different scanning path will result in a lower number, and therefore in a more economical identification process.

Going back to the code mobile it may be noted that, starting with a single chain, one additional subchain is created by each one of the nodes. Since the total number of links in the final (lowest) level is always  $P$ , it follows that the number of nodes is in all cases  $P - 1$ .

The way in which these  $P - 1$  nodes are distributed in the various levels is a determining factor in the amount of redundancy inherent in the given set and scanning path. The longest possible scanning path occurs when the nodes are distributed in  $P - 1$  different levels, and the number of cells required for identification is then  $P - 1$ . Consequently the number of redundant cells is never smaller than  $C - P + 1$ . In the best possible case the nodes are distributed in successive powers of 2, and the number of cells required for identification is then  $\log_2 P$ . Certainly such a case is impossible unless  $\log_2 P$  is an integer.

Thus the node distribution, as determined via the code mobile or its equivalent schedule, not only reveals the redundant cells in any given case, but also enables us to estimate the efficiency associated with the specified scanning path and the advisability of introducing an alternative one.

#### CONCLUSION

It has been found that the code mobile is an extremely useful instrument for achieving introspection as well as deriving general relationships pertinent to pattern recognition processes.

The code schedule, the numerical counterpart of the mobile, proved to be a satisfactory tool for fast determination of redundancies in practical examples.

It is believed that both the "mobile" and the "schedule" may be helpful not only in problems involving *analysis* of given recognition systems, but also in problems involving *synthesis* of such systems (*e.g.*, finding the sampling sequence which yields the largest number of disposable cells).

Finally it may be mentioned that the methods and conclusions presented in this paper are not restricted to *patterns* only, but may be directly extended to any class of "two-tone" storable messages.

#### ACKNOWLEDGMENT

The author wishes to express his thanks to T. F. Jones of the electrical engineering department of Massachusetts Institute of Technology for his encouragement.



## Correction

Kent R. Johnson, author of the article "Optimum, Linear, Discrete Filtering of Signals Containing a Non-random Component," which appeared on pages 49-55 of June, 1956 issue of IRE TRANSACTIONS ON INFORMATION THEORY, has requested that the following corrections be made by the replacements listed below.

"Predicted" with "predicated" in the second sentence of the Summary.

" $w_1$ " with " $w_i$ " in (11).

" $\psi$ " with " $\Psi$ " in (20b).

"B" with "E" on the left side of (23).

"N" with "H" on the left side of (25).

"Crammer's" with "Cramer's" in the sentence following (32).

" $p + 1$ " with " $p = 1$ " as the lower limit on the summation in the  $(1, k + 1)$  element of the determinant (34).

" $(N^2 + 1)$ " with " $(N^2 - 1)$ " in both (37) and (38).

The second minus sign on the right side of (45) with and equality sign.

Place brackets around the right side of (28).



## PGIT News

*(The following account of the 1956 Symposium on Information Theory is due to Prof. Peter Elias, chairman of the organizing committee, and does not note the tremendous effort expended by the organizing committee, particularly Prof. Elias, to whom much of the credit for the great success of this Symposium belongs.*

—The Editor

### 1956 SYMPOSIUM ON INFORMATION THEORY

The 1956 Symposium on Information Theory was held September 10-12 at the Kresge Auditorium of the Massachusetts Institute of Technology, Cambridge, Mass. Like the 1954 Symposium, it was organized by the Professional Group on Information Theory of the Institute of Radio Engineers and the Research Laboratory of Electronics of M.I.T. The Office of Naval Research, the Air Research and Development Command, the Signal Corps Engineering Laboratories, and URSI were cosponsors. The meeting was attended by about 300 people. It was more international than the 1954 meeting. It included two French scientists, Prof. B. Mandelbrot, of the University of Geneva, who gave a paper on thermodynamics, and Dr. M. P. Schutzenberger, now a visiting research associate at M.I.T., who gave a paper on coding. Two German scientists, Dr. T. von Randow, who is a Commonwealth Fund Fellow visiting at M.I.T.,

and Dr. H. Mittelstaedt, who is spending the year at Tufts University, attended. Three Russian scientists, Dr. B. V. Gnedenko, mathematician from the Mathematical Institute in Kiev, Dr. V. I. Siforov, chairman of the Russian equivalent of IRE, and Dr. D. Panov, director of the Institute of International Information, Academy of Sciences, USSR, were also at the meeting.

The nineteen papers scheduled for presentation at the meeting were printed in the September issue of the PGIT TRANSACTIONS. In addition, a paper by Dr. Siforov arrived Friday, September 7, in the mail, and a paper by A. N. Kolmogorov, the most distinguished living Russian mathematician, arrived on Sunday, September 9, with Dr. Gnedenko. These papers were translated and duplicated, and 300 copies were passed out to the Symposium participants on Tuesday morning, September 11. This made some intelligent discussion possible when the papers were presented on Wednesday. The PGIT is deeply indebted to Morris D. Friedman, who translated, and to *Electronics Translations*, who did the duplicating, for making this material available so rapidly. The translations are being published in this issue of the PGIT TRANSACTIONS to complete the permanent record of the Symposium.

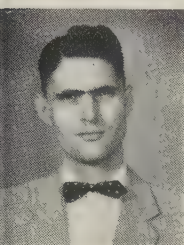
Some of the events at the Symposium do not appear in the TRANSACTIONS. One of these was a warm welcome and a de-

scription of the place of information theory at M.I.T., by Chancellor Stratton. A second was an informal report on Monday evening by Prof. J. McCarthy, of Dartmouth College, Dr. M. Minsky, of Harvard, and Dr. R. J. Solomonoff, of the Technical Research Center, of the results of a group which met at Dartmouth this Summer to study problems of artificial intelligence. A third was Prof. Norbert Wiener's after-dinner talk at the Tuesday banquet, on information theory and its relation to physics and the analysis of causality, which was enjoyed by an audience of 220. A fourth was Dr. Gnedenko's brief speech replying to the comments on Kolmogorov's paper. Gnedenko said that Russian mathematicians, as well as engineers, felt that information theory was the most important scientific event of the past decade.

The smooth running of the Symposium was due largely to its Treasurer, Ralph Sayers, who was in charge of arrangements, with the assistance of Mrs. N. Eidelberg, Miss D. Scalon, R. Keyes, and other members of the R.L.E. staff. The very considerable effort involved in making the Russian participation possible and providing translation service at the meeting was undertaken by the Hospitality Committee consisting of Dr. Paul Green, Prof. V. Yngve and Prof. M. Halle, of M.I.T., and Prof. L. Dolansky, of Northeastern. Dr. E. C. Rowan, of the National Academy of Sciences, was also of great assistance in helping with foreign participation.

# Contributors

L. Lorne Campbell (M '56) was born on October 20, 1928, in Winnipeg, Canada. He received the B.Sc. degree in mathematics and physics from the University of Manitoba in 1950.

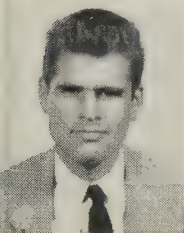


L. L. CAMPBELL

In 1951, he was conferred the M.S. degree in physics from Iowa State College, and in 1955, the Ph.D. degree in mathematics from the University of Toronto. Since 1954, Dr. Campbell has been employed at the Radio Physics Laboratory of the Defence Research Board in Ottawa, Canada. He is a member of the American Mathematical Society.



Peter Elias (S '48—A '51—M '56) was born on November 26, 1923, in New Brunswick, N. J. He received the B.S. degree from Massachusetts Institute of Technology, Cambridge, Mass., in 1944.



P. ELIAS

Dr. Elias then served with the U. S. Navy until 1946. In 1946 he entered Harvard University to do graduate work in digital computers and communications. He received the M.A. degree in 1948, the M.E.S. degree in 1949 and the Ph.D. degree in 1950 from Harvard. From 1950 to 1953, he was a Lowell Prize Fellow at Harvard, doing postdoctoral research on information theory.

In 1953 he was appointed assistant professor of electrical engineering at M.I.T. He has since been doing research on picture transmission and on coding problems in information theory at the Research Laboratory of Electronics, and teaching in the electrical engineering department, of which he is now an associate professor.

He is a member of Sigma Xi and of the Institute of Mathematical Statistics.



Amiel Feinstein was born in New York, N. Y., on January 25, 1930. He received the B.S. degree from Brooklyn College in 1950, and was conferred the Ph.D. degree in physics by Massachusetts Institute of Technology, Cambridge, Mass. His doctoral thesis was entitled "A New Basic Theorem of Information Theory."



A. FEINSTEIN

In the summer of 1954 he worked with the mathematics

group at the Bell Telephone Laboratories in Murray Hill, N. J.

Since that time, Dr. Feinstein has been employed by the Lincoln Laboratory of M.I.T. at Lexington, Mass. He has been working on statistical problems in communications and radar.



Lawrence J. Fogel (A '49—M '53) was born in Brooklyn, N. Y., on March 2, 1928. He received the B.E.E. degree from New York University in June, 1948, the M.S. in E.E. degree from Rutgers University in June, 1952, and is currently working on the dissertation for his doctorate degree at Polytechnic Institute of Brooklyn.



L. J. FOGEL

From June, 1948, through June, 1953, he was employed at Fort Monmouth, N. J., primarily in connection with the electronic aspects of the Army Aviation Program. This work included antenna design and evaluation.

Mr. Fogel was then employed as staff engineer by Stavid Engineering, Inc., Plainfield, N. J., where he initiated a program of human engineering using an application of communication and servo theory. For a period of time he was associated, in a consultant capacity, with the Psychology Branch of the Aero-Medical Laboratory, Wright Air Development Center.

Currently Mr. Fogel is employed as staff engineer and design specialist at Convair, San Diego, Calif., where he is concerned with systems analysis and human engineering for cockpit control-display design.



Arthur Glovazky (S '53) was born in Haifa, Israel, on April 18, 1930. In 1950 he entered Massachusetts Institute of Technology, Cambridge, Mass., where in 1955 he received the B.S. degree and in 1956 the M.S. degree in electrical engineering.



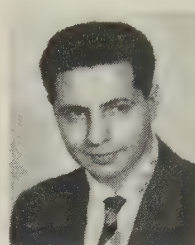
A. GLOVAZKY

From 1954 to 1956 he served as a teaching assistant in the department of electrical engineering at M.I.T.; in this capacity he instructed undergraduate laboratories and elementary courses in network analysis.

At present, he is employed by the Raytheon Manufacturing Company, Waltham, Mass., as a research engineer.

Mr. Glovazky is a member of the American Institute of Electrical Engineers, Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

David L. Jagerman (S '49—A '50—M '53) was born in New York, N. Y., on August 27, 1923. He received the B.E.E. degree in 1949 from Cooper Union, the M.S. degree in 1954 from New York University, and is currently working on his dissertation for the doctoral degree in applied mathematics, also from New York University.



D. L. JAGERMAN

From 1943 to 1945, he served with the U. S. Army Signal Corps, where he was involved in the operation and maintenance of aircraft communications equipment.

From 1949 to 1950, he served as test engineer for United Transformer Co., and from 1950 to 1951, he was employed at Evans Signal Laboratory as design engineer. He was engineer and mathematician for Project Cyclone at Reeves Instrument Corp. from 1951 until 1955, when he joined the engineering staff of Stavid Engineering, Inc., Plainfield, N. J., as senior engineer, where he is currently employed.

He has been especially active in the fields of analog computation, noise analyses, the mathematical analyses of guided missile systems, and the application of interpolation theory to the investigation of physical systems.



Andrei N. Kolmogorov was born on April 25, 1903. He entered Moscow University in 1920 and studied mathematical



A. N. KOLMOGOROV

research. During the next ten years he published several contributions on trigonometric series, generalized differentiation, measure theory, mathematical logic, integration, functional analysis, and topology.

In 1927 he published a formulation of an axiomatic basis for probability theory. Professor Kolmogorov has also engaged in studies of applications of probability theory to the study of Brownian motion, diffusion, crystallization, quality-control, meteorological forecasting, and fire-control, and has formulated theories of turbulence and statistical interpolation and extrapolation.

He received the Stalin Prize from the Soviet government in 1940. In 1939 he was elected Academician of the Academy of Sciences of the U.S.S.R. He is Director of the Institute of Mathematics and Mechanics of Moscow State University, and Chairman of the Division of Probability Theory and Mathematical Statistics of the Steklov Mathematical Institute.

For a photograph and biography of J. A. McFadden, see p. 97 of the June, 1956, issue of IRE TRANSACTIONS ON INFORMATION THEORY.



Brockway McMillan (SM '54) was born on March 30, 1915 in Minneapolis, Minn. He attended Armour Institute, Chicago, Ill., from 1932 until 1934, when he entered M.I.T., where he received the B.S. degree in 1936 and the Ph.D. degree in 1939, both in mathematics. He was a part-time instructor at M.I.T. during 1936-1939.



B. J. McMILLAN

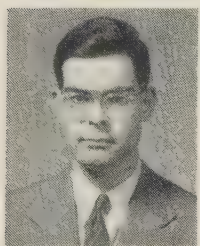
From 1939 to 1943, he was a Proctor Fellow, then Fine Instructor, and later research associate at Princeton University. He was on active duty with the Naval Reserves during World War II.

Dr. McMillan joined the staff of Bell Telephone Labs., Inc., in 1946, as a research mathematician working principally with network theory and the statistical and mathematical theory of communication. For the past year, he has been assistant director of systems engineering.

He is a member of the American Mathematical Society, Institute of Mathematical Statistics, Society for Industrial and Applied Mathematics, and the American Association for the Advancement of Science.



Robert Price (S'48-A'54) was born in West Chester, Pa., on July 7, 1929. He received the A.B. degree in physics from Princeton University in 1950, and the Sc.D. degree in electrical engineering from the Massachusetts Institute of Technology in 1953.



R. PRICE

While at Princeton, he participated in the 1949 cosmic ray expedition to the Central Pacific and during the summers was employed in tele-

vision research at the Philco Corporation. At M.I.T. he held a fellowship in the Research Laboratory of Electronics and later transferred to the Lincoln Laboratory where he carried out his doctoral thesis on communication via multipath. For a year he was engaged in radio-astronomy under a Fulbright award at the Commonwealth Scientific and Industrial Research Organization in Sydney, Australia. Since 1954 he has been a staff member of the Lincoln Laboratory.

Dr. Price is a member of Phi Beta Kappa, Sigma Xi, and the Franklin Institute.



For a photograph and biography of Mischa Schwartz, see p. 98 of the June, 1956, issue of IRE TRANSACTIONS ON INFORMATION THEORY.



Claude E. Shannon (S '36—M '48—SM '49—F '50) was born in Gaylord, Mich., on April 30, 1916. He received the bachelor's degree in electrical engineering and mathematics from the University of Michigan. After four years of graduate study at M.I.T., he was awarded the master's degree in electrical engineering and the Ph.D. degree in mathematics in 1940.



C. E. SHANNON

During these years at M.I.T. Dr. Shannon served as research assistant in the electrical engineering department and later as an assistant in the mathematics department.

As a National Research Fellow Dr. Shannon studied at the Institute for Advanced Study, Princeton, N.J., in 1940, and in 1941 he joined the staff of the Bell Telephone Laboratories, Murray Hill, N.J., as a research mathematician, and earlier this year, he was appointed visiting professor of electrical communications at M.I.T.

Dr. Shannon has contributed to many fields of applied mathematics. At M.I.T., he was in charge of operating the Institute's differential analyzer. His work in mathematical theory has also led to the de-

velopment of maze-solving machines, and he has more recently undertaken such probability problems as the design of reliable machines consisting of unreliable components.

Dr. Shannon's work has been recognized by the award of the Alfred Nobel Prize of the American Institute of Electrical Engineers, the Morris Liebmann Award of the IRE, and the Stuart Ballantine Medal of the Franklin Institute. In 1954 he was awarded the honorary degree of Master of Science by Yale University.

Dr. Shannon is a member of Sigma Xi, Phi Kappa Phi, and the American Mathematical Society. He is co-author of a book on the mathematical theory of communication, the author of a number of technical papers, and the holder of a number of patents.



Vladimir I. Siforov was born May 31, 1904 in Moscow. He entered the Mechanical-Electro-Engineering Institute in Moscow and then attended the Electro-Engineering Institute in Leningrad, graduating in the radio faculty in 1929.



V. I. SIFOROV

In 1927 he began working in various factories and research institutes of the radio field. Beginning in 1930 he was lecturer on the theory of alter-

nating currents and radio receivers in the Electro-Engineering Institute.

He was awarded the degree of doctor of technical sciences in 1937, and in 1938 was awarded the degree of professor in the chair of radio receivers, a position he has held in several Leningrad universities.

Since 1953 Dr. Siforov has been the director of the Research Institute of the Ministry of Communications, USSR. The same year he was elected associate member of the Academy of Sciences.

Since 1954 he has been in charge of a laboratory in the Research Radio and Electronic Engineering Institute of the Soviet Academy of Sciences. He is chairman of the Central Board of the Scientific Radio and Electrical Society.



Index to

IRE TRANSACTIONS

ON

INFORMATION THEORY

Volume IT-2, 1956

# IRE Transactions on Information Theory

## Index to Volume IT-2—1956

Volume IT-2, Number 1, March, 1956		Index Number	Page
Index Number			
	Preliminary Announcement—Symposium on Information Theory . . . . .		1
	Annual ACM Meeting . . . . .		1
	Frontispiece, <i>C. E. Shannon</i> . . . . .		2
	The Bandwagon, <i>C. E. Shannon</i> . . . . .		3
IT94.	The Probability Distribution for the Filtered Output of a Multiplier, <i>D. G. Lampard</i> . . . . .		4
IT95.	Interpolation and Extrapolation of Sampled Data, <i>A. B. Lees</i> . . . . .		12
IT96.	Report on London Symposium, <i>N. M. Blachman</i> . . . . .		17
IT97.	Some Optimal Signals for Time Measurements, <i>H. Sherman</i> . . . . .		24
IT98.	An Analysis of Signal Detection and Location by Digital Methods, <i>G. P. Dinneen and I. S. Reed</i> . . . . .		29
IT99.	Inverse Probability in Angular Tracking Radars, <i>B. J. DuWaldt</i> . . . . .		38
	Correspondence:		
IT100.	The Linear, Input-Controlled Variable-Pass Network, by <i>B. E. Keiser, A. Jeffrey</i> . . . . .		42
IT101.	Rebuttal, <i>B. E. Keiser</i> . . . . .		43
	Contributors . . . . .		44
Volume IT-2, Number 2, June, 1956			
	Frontispiece, <i>N. Wiener</i> . . . . .		47
	What Is Information Theory? <i>N. Wiener</i> . . . . .		48
IT102.	Optimum, Linear, Discrete Filtering of Signals Containing a Nonrandom Component, <i>K. R. Johnson</i> . . . . .		49
IT103.	Spatial Filtering in Optics, <i>E. L. O'Neill</i> . . . . .		56
IT104.	Effects of Signal Fluctuation on the Detection of Pulse Signals in Noise, <i>M. Schwartz</i> . . . . .		66
IT105.	Solution of an Integral Equation Occurring in the Theories of Prediction and Detection, <i>K. S. Miller and L. A. Zadeh</i> . . . . .		72
IT106.	Generalization of the Class of Nonrandom Inputs of the Zadeh-Ragazzini Prediction Model, <i>M. Blum</i> . . . . .		76
IT107.	The Correlation Function of a Sine Wave Plus Noise After Extreme Clipping, <i>J. A. McFadden</i> . . . . .		82
IT108.	A Note on Two Binary Signaling Alphabets, <i>D. Stepan</i> . . . . .		84
IT109.	Generating a Gaussian Sample, <i>S. Stein and J. E. Storer</i> . . . . .		87
IT110.	A Bibliography of Soviet Literature on Noise, Correlation, and Information Theory, <i>P. E. Green, Jr.</i> . . . . .		91
IT111.	On the Information Invariant, <i>S. Okada</i> . . . . .		95
	Correspondence:		
IT112.	Letters on "In Which Fields Do We Graze?," <i>M. Hoberman, S. Bagno, and N. M. Blachman</i> . . . . .		96
	Contributors . . . . .		97
Volume IT-2, Number 3, September, 1956			
1956 Symposium on Information Theory			
IT113.	The Zero Error Capacity of a Noisy Channel, <i>C. E. Shannon</i> . . . . .		S8
IT114.	A Linear Circuit Viewpoint on Error-Correcting Codes, <i>D. A. Huffman</i> . . . . .		S20
IT115.	Theory of Information Feedback Systems, <i>S. S. L. Chang</i> . . . . .		S29
IT116.	A Linear Coding for Transmitting a Set of Correlated Signals, <i>H. P. Kramer and M. V. Mathews</i> . . . . .		S41
	IT117. On an Application of Semi-Groups Methods To Some Problems in Coding, <i>M. P. Schutzenberger</i> . . . . .		S4
	IT118. The Logic Theory Machine: A Complex Information Processing System, <i>A. Newell and H. A. Simon</i> . . . . .		S6
	IT119. Tests on a Cell Assembly Theory of the Action of the Brain, Using a Large Digital Computer, <i>N. Rochester, J. H. Holland, L. H. Haibt, and W. L. Duda</i> . . . . .		S8
	IT120. The Measurement of Third Order Probability Distributions of Television Signals, <i>W. F. Schreiber</i> . . . . .		S9
	IT121. Gap Analysis and Syntax, <i>V. H. Yngve</i> . . . . .		S10
	IT122. Three Models for the Description of Language, <i>N. Chomsky</i> . . . . .		S11
	IT123. Some Studies in the Speed of Visual Perception, <i>G. C. Sziklai</i> . . . . .		S12
	IT124. Human Memory and the Storage of Information, <i>G. A. Miller</i> . . . . .		S12
	IT125. The Human Use of Information III. Decision-Making in Signal Detection and Recognition Situations Involving Multiple Alternatives, <i>J. A. Swets and T. G. Birdsall</i> . . . . .		S13
	IT126. On Optimum Nonlinear Extraction and Coding Filters, <i>A. V. Balakrishnan and R. Drenick</i> . . . . .		S16
	IT127. Final-Value Systems with Gaussian Inputs, <i>R. C. Booton, Jr.</i> . . . . .		S17
	IT128. An Extension of the Minimum Mean Square Prediction Theory for Sampled Input Signals, <i>M. Blum</i> . . . . .		S17
	IT129. A New Interpretation of Information Rate, <i>J. L. Kelly, Jr.</i> . . . . .		S18
	IT130. An Outline of a Purely Phenomenological Theory of Statistical Thermodynamics: I. Canonical Ensembles, <i>B. Mandelbrot</i> . . . . .		S19
	IT131. A Radar Detection Philosophy, <i>W. McC. Siebert</i> . . . . .		S20
Volume IT-2, Number 4, December, 1956			
	Frontispiece, <i>M. J. Di Toro, Jr.</i> . . . . .		10
	Editorial, <i>M. J. Di Toro, Jr.</i> . . . . .		10
	IT132. On the Shannon Theory of Information Transmission in the Case of Continuous Signals, <i>A. N. Kolmogorov</i> . . . . .		10
	IT133. On Noise Stability of a System with Error-Correcting Codes, <i>V. I. Siforov</i> . . . . .		10
	IT134. Two Inequalities Implied by Unique Decipherability, <i>B. McMillan</i> . . . . .		11
	IT135. A Note on Maximum Flow Through a Network, <i>P. Elias, A. Feinstein, and C. E. Shannon</i> . . . . .		11
	IT136. Rectification of Two Signals in Random Noise, <i>L. L. Campbell</i> . . . . .		11
	IT137. Optimum Detection of Band-Pass Gaussian Signals in White Gaussian Noise, with Application to Receiver Design for Scatter-Path Communication, Part I, <i>R. Price</i> . . . . .		12
	IT138. A Coincidence Procedure for Signal Detection, <i>M. Schwartz</i> . . . . .		13
	IT139. Some General Aspects of the Sampling Theorem, <i>D. L. Jagerman and L. J. Fogel</i> . . . . .		13
	IT140. The Axis-Crossing Intervals of Random Functions, <i>J. A. McFadden</i> . . . . .		14
	IT141. Determination of Redundancies in a Set of Patterns, <i>A. Glovazky</i> . . . . .		15
	IT142. Correction to "Optimum, Linear, Discrete Filtering of Signals Containing a Nonrandom Component," <i>K. R. Johnson</i> . . . . .		15
	PGIT News . . . . .		15
	Contributors . . . . .		15

# Index to Authors

- |  |   |   |   |
|--|---|---|---|
| <p><b>B</b></p> <p>Bagno, S.: IT112<br/>         Balakrishnan, A. V.: IT126<br/>         Birdsall, T. G.: IT125<br/>         Blachman, N. M.: IT96,<br/>             IT112<br/>         Blum, M.: IT106, IT128<br/>         Booton, R. C., Jr.: IT127</p> <p><b>C</b></p> <p>Campbell, L. Lorne: IT136<br/>         Chang, S. S. L.: IT115<br/>         Chomsky, A.: IT122</p> <p><b>D</b></p> <p>Dinneen, G. P.: IT98<br/>         Drenick, R.: IT126<br/>         Duda, W. L.: IT119<br/>         DuWaldt, B. J.: IT99</p> <p><b>E</b></p> <p>Elias, P.: IT135</p> | <p><b>F</b></p> <p>Feinstein, A.: IT135<br/>         Fogel, L. J.: IT139</p> <p><b>G</b></p> <p>Glovazky, A.: IT141<br/>         Green, P. E., Jr.: IT110</p> <p><b>H</b></p> <p>Haibt, L. H.: IT119<br/>         Hoberman, M.: IT112<br/>         Holland, J. H.: IT119<br/>         Huffman, D. A.: IT114</p> <p><b>J</b></p> <p>Jagerman, D. L.: IT139<br/>         Jeffrey, A.: IT100<br/>         Johnson, K. R.: IT102, IT142</p> <p><b>K</b></p> <p>Keiser, B. E.: IT101<br/>         Kelly, J. L., Jr.: IT129<br/>         Kolmogorov, A. N.: IT132<br/>         Kramer, H. P.: IT116</p> | <p><b>L</b></p> <p>Lampard, D. G.: IT94<br/>         Lees, A. B.: IT95</p> <p><b>M</b></p> <p>Mandelbrot, B.: IT130<br/>         Mathews, M. V.: IT116<br/>         McFadden, J. A.: IT107,<br/>             IT140<br/>         McMillan, B.: IT115<br/>         Miller, G. A.: IT124<br/>         Miller, K. S.: IT105</p> <p><b>N</b></p> <p>Newell, A.: IT118</p> <p><b>O</b></p> <p>Okada, S.: IT111<br/>         O'Neill, E. L.: IT103</p> <p><b>P</b></p> <p>Price, R.: IT137</p> | <p><b>R</b></p> <p>Reed, I. S.: IT98<br/>         Rochester, N.: IT119</p> <p><b>S</b></p> <p>Schreiber, W. F.: IT120<br/>         Schutzenberger, M. P.: IT117<br/>         Schwartz, M.: IT104, IT138<br/>         Shannon, C. E.: IT113, IT135<br/>         Sherman, H.: IT97<br/>         Siebert, W. McC.: IT131<br/>         Siforov, V. I.: IT133<br/>         Simon, H. A.: IT118<br/>         Slepian, D.: IT108<br/>         Stein, S.: IT109<br/>         Storer, J. E.: IT109<br/>         Swets, J. A.: IT125<br/>         Sziklai, G. C.: IT123</p> <p><b>Y</b></p> <p>Yngve, V. H.: IT121</p> <p><b>Z</b></p> <p>Zadeh, L. A.: IT105</p> |
|--|---|---|---|

# Index to Technical Subjects

- |   |   |
|---|---|
| <p><b>A</b></p> <p>Alphabets, Binary Signaling: IT108<br/>         Axis-Crossing Intervals of Random Functions: IT140</p> <p><b>B</b></p> <p>Bandwidth Reduction Using Nonlinear Encoding and Decoding<br/>             Filters: IT126<br/>         Binary Signaling Alphabets: IT108<br/>         Brain, Cell Assembly Theory of Action Tested by Digital Com-<br/>             puter: IT119</p> <p><b>C</b></p> <p>Cell Assembly Theory of Brain Action Tested by Digital Com-<br/>             puter: IT119<br/>         Channel Capacity, Zero Error Capacity of Noisy Channel: IT113<br/>         Coding: IT113, IT114, IT116, IT117, IT126, IT133<br/>             Error-Correcting Codes: IT114, IT133<br/>             Linear Circuit Viewpoint for: IT114<br/>             Noise Stability of: IT133<br/>         Filters and Extraction Filters: IT126<br/>         Semi-Group Methods Applied: IT117<br/>         for Transmission of Correlated Signals: IT116<br/>         Zero Error Capacity of Noisy Channel IT113<br/>         Coincidence Procedure for Signal Detection: IT138<br/>         Computers, Digital, in Tests on Cell Assembly Theory of Action<br/>             of Brain: IT119<br/>         Correlation: IT107, IT110<br/>             Function of a Sine Wave Plus Noise: IT107<br/>             Soviet Literature on: IT110</p> <p><b>D</b></p> <p>Data Handling Systems, Logic Theory Machine: IT118<br/>         Decipherability, Unique, Inequalities Implied by: IT134<br/>         Decision-Making in Signal Detection and Recognition Situations:<br/>             IT125<br/>         Detection: IT104, IT105, IT125, IT131, IT137, IT138<br/>             Coincidence Procedure for: IT138</p> | <p>Gaussian Signals in Noise for Scatter-Path Communication:<br/>             IT137<br/>         Pulse, Effects of Signal Fluctuation: IT104<br/>         Radar, Philosophical Approach to Problems: IT131<br/>         and Recognition of Signals by Human Observers: IT125<br/>         Solution of an Integral Equation: IT105<br/>         Digital Methods of Signal Detection and Location: IT98</p> <p><b>E</b></p> <p>Error-Correcting System, Noise Stability of: IT133<br/>         Extraction Filters and Coding Filters: IT126</p> <p><b>F</b></p> <p>Feedbacks, Information Theory of: IT115<br/>         Filtering: IT102, IT103<br/>             of Signals Containing a Nonrandom Component: IT102<br/>             Spatial, in Optics: IT103<br/>         Filters: IT114, IT126<br/>             Binary, Linear Circuit Viewpoint on Error-Correcting Codes:<br/>                 IT114<br/>             Optimum Nonlinear Extraction and Coding: IT126<br/>         Final-Value Systems with Gaussian Inputs: IT127</p> <p><b>G</b></p> <p>Gap Analysis and Syntax: IT121<br/>         Gaussian Sample, Generating of: IT109<br/>         Generating a Gaussian Sample: IT109</p> <p><b>H</b></p> <p>Human Memory and Storage of Information: IT124<br/>         Human Use of Information, Decision-Making: IT125</p> <p><b>I</b></p> <p>In Which Fields Do We Graze?: IT112<br/>         Information Feedback Systems, Theory of: IT115<br/>         Information Invariant: IT111<br/>         Information Rate, New Interpretation of: IT129</p> |
|---|---|

Information Storage and Human Memory: IT124  
 Information Theory: IT96, IT110, IT112  
 Applications of: IT112  
   London Symposium: IT96  
   Soviet Literature on: IT110  
 Information Transmission of Continuous Signals: IT132  
 Inverse Probability in Angular-Tracking Radars: IT99

## L

Language: IT121, IT122  
   Descriptive Models: IT122  
   Gap Analysis and Syntax: IT121  
 Logic Theory Machine: IT118  
 London Symposium on Information Theory: IT96

## M

Mean Square Operations: IT126, IT127, IT128  
   Final Value Systems with Gaussian Inputs: IT127  
   Optimum Nonlinear Extraction and Coding Filters: IT126  
   Prediction Theory for Sampled Input Signals: IT128  
 Measurements: IT97, IT120  
   of Third Order Probability Distributions of Video Signals:  
     IT120  
   Time, Optimal Signals for: IT97  
 Memory and Storage of Information by Humans: IT124  
 Multiplier Output, Probability Distribution for: IT94

## N

Networks: IT100, IT101, IT135  
   Input-Controlled, Variable-Pass: IT100, IT101  
   Maximum Flow Through: IT135  
 Noise: IT98, IT102, IT107, IT109, IT110, IT113, IT133, IT136,  
   IT137  
   Detection of Band-Pass Signals for Scatter-Path Communi-  
     cation: IT137  
   Filtering of Nonrandom Signals: IT102  
   Gaussian: IT107, IT109  
     Generation of: IT109  
     and a Sine Wave Correlation Function: IT107  
   Rectification of Two Signals in: IT136  
   Signal Detection and Location by Digital Methods: IT98  
   Soviet Literature on: IT110  
   Stability of Error-Correcting Systems: IT133  
   Zero Error Capacity of Noisy Channel: IT113  
 Nonrandom Inputs of Zadeh-Ragazzini Prediction Model: IT106

## O

Optics, Spatial Filtering: IT103

## P

Pattern Redundancies, Determination of: IT141  
 Phenomenological Theory of Statistical Thermodynamics: IT130  
 Prediction: IT105, IT106, IT128  
   Nonrandom Inputs of Zadeh-Ragazzini Model: IT106  
   Solution of an Integral Equation: IT105  
   Theory for Sampled Input Signals, Minimization of Mean  
     Square Error: IT128

Probability Distribution, Third Order, Measurement in Television  
 Signals: IT120

Probability Theory, New Interpretation of Information Rate: IT12  
 Pulse Detection, Effects of Signal Fluctuation: IT104

## R

Radar: IT99, IT131  
   Angular-Tracking, Inverse Probability in: IT99  
   Detection Philosophy: IT131  
 Random Functions, Axis-Crossing Intervals of: IT140  
 Receivers: IT137  
   Signal in Noise Detection for Scatter-Path Communication  
     IT137  
 Recognition and Detection of Signals by Human Observers: IT125  
 Rectification of Two Signals in Noise: IT136  
 Redundancies in Set of Patterns, Determination of: IT141

## S

Sampled Data, Interpolation and Extrapolation: IT95  
 Sampling Theorem, General Aspects of: IT139  
 Scatter-Path Communication Receivers, Detection in Noise: IT13  
 Semi-Group Methods Applied to Coding Problems: IT117  
 Shannon Information Transmission Theory for Continuous Signals  
   IT132  
 Signaling Alphabets, Binary: IT108  
 Signals: IT97, IT98, IT102, IT104, IT106, IT107, IT109, IT116,  
   IT125  
   Correlated, Linear Coding for Transmission: IT116  
   Correlation Function of a Sine Wave Plus Noise: IT107  
   Detection and Location by Digital Methods: IT98  
   Detection and Recognition by Human Observers: IT125  
   Effects of Fluctuations on Pulse Detection: IT104  
   Generating a Gaussian Sample: IT109  
   Nonrandom Filtering of: IT102  
   Nonrandom Inputs of the Zadeh-Ragazzini Prediction Model:  
     IT106  
     for Time Measurement: IT97  
 Soviet Literature on Noise, Correlation and Information Theory  
   IT110  
 Spatial Filtering in Optics: IT103  
 Storage of Information and Memory in Humans: IT124  
 Symposium on Information Theory, 1956: IT113-131  
 Syntax and Gap Analysis: IT121

## T

Television: IT120, IT123  
   Signals, Measurement of Third Order Probability Distribution  
     IT120  
   Visual Perception Studies: IT123  
 Thermodynamics, Statistical, Phenomenological Theory of: IT130  
 Time Measurements, Optimal Signals, for: IT97

## V

Visual Perception, Studies in Speed of: IT123

## Z

Zero Error Capacity of Noisy Channel: IT113

# Nontechnical Index

## Announcements

Association for Computing Machinery  
   Annual Meeting: March, p. 1  
 Information Theory Symposium; March, p. 1, December, p. 154

## Biographies of Contributors

March, p. 44  
 June, p. 97  
 December, p. 155

## Editorials

"The Bandwagon," by Claude E. Shannon: March, p. 3  
 "What Is Information Theory?" by Norbert Wiener: June, p. 48  
 "Application of Information Theory," by Michael J. Di Toro  
   December, 1956

## Frontispiece

Shannon, Claude E.: March, p. 2  
 Wiener, Norbert: June, p. 47  
 Di Toro, Michael J., Jr.: December, p. 100







## INFORMATION FOR AUTHORS



Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT TRANSACTIONS. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in india ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the PROCEEDINGS OF THE IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE TRANSACTIONS ON INFORMATION THEORY is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. A period of approximately two months additional is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. In addition, the IRE CONVENTION RECORD is published in July, and a bound collection of Information Theory papers delivered at the annual IRE National Convention is mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to Laurin G. Fischer, Federal Telecommunication Labs., 492 River Road, Nutley, N. J. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Nathan Marchand, Marchand Electronic Labs., Riversville Road, Greenwich, Conn.